

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: November 26, 2005, 12:53:47 ; Search time 8121.42 Seconds
(without alignment)
10694.773 Million cell updates/sec

Title: US-09-555-529-1
Perfect score: 1528
Sequence: 1 ctgaattcagccgcgctg.....atgtattataaaaaaa 1528

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2006000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_in.*
- 3: gb_env.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pri.*
- 9: gb_ro.*
- 10: gb_sts.*
- 11: gb_sv.*
- 12: gb_un.*
- 13: gb_vl.*
- 14: gb_hg.*
- 15: gb_pl.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1528	100.0	1528	6	AX003308	Sequence
2	1528	100.0	1528	8	HSJ5273	Sequence
3	1453	95.1	2059	8	BC017309	Homo sapi
4	1002	65.6	1002	6	AX003311	Sequence
5	999.8	65.4	1414	9	MMK117	Mouse KIN17
6	996.4	65.2	1366	9	BC028860	Mus muscu
7	990.6	64.8	1390	6	AX003331	Sequence
8	969.2	63.4	221547	14	AC119716	AC119716 Rattus no
9	955.2	62.5	1117	6	CQ722054	Sequence
10	759	49.7	130262	14	AC166154	AC166154 Mus muscu
11	759	49.7	140215	9	AC140205	AC140205 Mus muscu
12	705	46.1	1483	5	BC074606	BC074606 Xenopus t
13	685	44.8	144118	14	AC141995	AC141995 Rattus no
14	673.2	44.1	180882	14	AC114615	AC114615 Mus muscu
15	673.2	44.1	182400	14	AC124010	AC124010 Mus muscu
16	673.2	44.1	185378	9	AL772367	AL772367 Mouse DNA
17	672.4	44.0	1501	5	BC079988	BC079988 Xenopus l
18	663.8	43.4	2710	9	BC058169	BC058169 Mus muscu

19	653.4	42.8	1002	6	AX003310	Sequence	
20	475	31.1	1102	6	AX003309	Sequence	
21	465.4	30.5	175480	14	AC127074	Rattus no	
22	465.4	30.5	247615	14	AC098061	Rattus no	
23	436.4	28.6	461	6	CQ684251	Sequence	
C	24	402.8	26.4	177276	14	AC154247	Mus muscu
25	380.6	24.9	1372	2	AK113340	Ciona int	
C	26	337	22.1	174361	8	AL158044	Human DNA
C	27	337	22.1	194563	14	AL391687	Homo sapi
28	333	21.8	333	6	AX003340	Sequence	
29	300.2	19.6	1467	2	AY084184	Drosophil	
30	298.6	19.5	1241	6	CQ613281	Sequence	
31	284	18.6	300	6	BD220437	Human gen	
32	261.6	17.1	333	6	AX003341	Sequence	
33	249.6	16.3	616	10	BV448133	KIN_9291	
34	247.2	16.2	3313	6	CQ613280	Sequence	
C	35	247.2	16.2	31807	14	AC017679	Drosophil
36	247.2	16.2	185918	2	AC009839	Drosophil	
37	247.2	16.2	301136	2	AE003591	Drosophil	
38	247	16.2	84798	14	AP007827	Lotus cor	
39	224.8	14.7	1563	15	AF360132	Arabidops	
C	40	224.8	14.7	87967	15	AC005223	Arabidops
41	224.4	14.7	1267	15	AY051011	Arabidops	
42	218	14.3	760	2	DME6528	Drosophil	
43	209.4	13.7	4394	6	CQ613259	Sequence	
44	209.4	13.7	4582	6	CQ596774	Sequence	
45	208.2	13.6	1824	15	AK069396	Oryza sat	

ALIGNMENTS

RESULT 1
AX003308 LOCUS
DEFINITION Sequence 1 from Patent WO9929845.
ACCESSION AX003308
VERSION AX003308.1 GI:9927125
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1
AUTHORS Angulo-Mora, J.F. and Mauffrey, P.
TITLES Sequences coding for kin17 protein and their applications
JOURNAL Patent: WO 9929845-A 1 17-JUN-1999;
ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE (FR)

FEATURES
source Location/Qualifiers
1..1528
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN	Query Match	100.0%;	Score 1528;	DB 6;	Length 1528;
	Best Local Similarity	100.0%;	Pred. No. 2e-311;		
	Matches 1528;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	CTAGATTTCAGCGCGCGCTGTAATTCCTAGAACTGGGTCAGAAAGTGATCGTCCGCTGG	60		
DB	1	CTAGATTTCAGCGCGCGCTGTAATTCCTAGAACTGGGTCAGAAAGTGATCGTCCGCTGG	60		
QY	61	TGCGCATGGGAGTGGGATTTCTTACTCCCAAGGCTATCGCCAAACAGATCAAGTCCA	120		
DB	61	TGCGCATGGGAGTGGGATTTCTTACTCCCAAGGCTATCGCCAAACAGATCAAGTCCA	120		
QY	121	AGGGGCTGCAGAAAGCTACCTGGTATTCAGATGTCAGAAAGTATCGCGGACGAGA	180		
DB	121	AGGGGCTGCAGAAAGCTACCTGGTATTCAGATGTCAGAAAGTATCGCGGACGAGA	180		

QY 181 ATGGCTTTAAGTGTCAATTTGTATGTCGGAATCTCATCAGAGACAACATATTGCTGGCTTCAG 240
Db 181 ATGGCTTTAAGTGTCAATTTGTATGTCGGAATCTCATCAGAGACAACATATTGCTGGCTTCAG 240
QY 241 AAAATCCTCAGCAGTTTATGGAATATTTTTCAGAGGAATTCGGAATGACTTTCTTAGAAC 300
Db 241 AAAATCCTCAGCAGTTTATGGAATATTTTTCAGAGGAATTCGGAATGACTTTCTTAGAAC 300
QY 301 TTCTCAGAGAGCGTTTGGCACTAAAGGGTCCACAACAACATTTGCTACACGAATACA 360
Db 301 TTCTCAGAGAGCGTTTGGCACTAAAGGGTCCACAACAACATTTGCTACACGAATACA 360
QY 361 TCAGCCACGAGAGACATCCATGAATGCCACTCAGTGGGAAACTCTGACTGATTTTA 420
Db 361 TCAGCCACGAGAGACATCCATGAATGCCACTCAGTGGGAAACTCTGACTGATTTTA 420
QY 421 CTAAGTGGCTGGCAGAGAGAGGCTTGTCAAAGTGGACGAGACACAAAAGGCTCGTATA 480
Db 421 CTAAGTGGCTGGCAGAGAGAGGCTTGTCAAAGTGGACGAGACACAAAAGGCTCGTATA 480
QY 481 TTCAGTACATAGACAGGACCCAGAAACTATCCCGCGCAACTGGAACTGGAGAAAAAGA 540
Db 481 TTCAGTACATAGACAGGACCCAGAAACTATCCCGCGCAACTGGAACTGGAGAAAAAGA 540
QY 541 AAAAGCAGGACTTCATGATGAAGAAAAAATGCCCCAAATTTATTGAAGAGCAAGTGAGAA 600
Db 541 AAAAGCAGGACTTCATGATGAAGAAAAAATGCCCCAAATTTATTGAAGAGCAAGTGAGAA 600
QY 601 GAGGCTCGAAGGGAAGAAACAGGAGGTCCTTACTTTTACGGAATTTAAGACGAGAAAAATG 660
Db 601 GAGGCTCGAAGGGAAGAAACAGGAGGTCCTTACTTTTACGGAATTTAAGACGAGAAAAATG 660
QY 661 ATGAAGAGAAAGTCAAGTTTAAATTTAGTAAAGGAGCATGTAGCTCATCCGAGCAACAT 720
Db 661 ATGAAGAGAAAGTCAAGTTTAAATTTAGTAAAGGAGCATGTAGCTCATCCGAGCAACAT 720
QY 721 CTTCCAGTCAAGTACTCTGGGACCGAGTGCATGAAGACGATAGGAATTCAGCATCAG 780
Db 721 CTTCCAGTCAAGTACTCTGGGACCGAGTGCATGAAGACGATAGGAATTCAGCATCAG 780
QY 781 TGAACAGAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAAAGAAAAAGAAAT 840
Db 781 TGAACAGAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAAAGAAAAAGAAAT 840
QY 841 CTGCACTGGATGAATCATGAGATTGAAGAGGAAAAAGAAAGAACTCCCGCAACAGACT 900
Db 841 CTGCACTGGATGAATCATGAGATTGAAGAGGAAAAAGAAAGAACTCCCGCAACAGACT 900
QY 901 ACTGGCTACAGCCTGAATTTATGTAAGGAAATTTATTAACCAAGAACTGGAGGAAATATC 960
Db 901 ACTGGCTACAGCCTGAATTTATGTAAGGAAATTTATTAACCAAGAACTGGAGGAAATATC 960
QY 961 ATAAGAAAAAGGCTATTGTTAAGGAAGTAATTTGACAAATATACAGCTGTTGGAAGATGA 1020
Db 961 ATAAGAAAAAGGCTATTGTTAAGGAAGTAATTTGACAAATATACAGCTGTTGGAAGATGA 1020
QY 1021 TTGATTCTGGAGACAAGCTGAAACTTGAACGACTCATTTAGAGACAGTAATTCAGACAC 1080
Db 1021 TTGATTCTGGAGACAAGCTGAAACTTGAACGACTCATTTAGAGACAGTAATTCAGACAC 1080
QY 1081 CAGGAAAAAGAAATCTAGTTTAAATGAGGCTACAGAGGAAATGAAGGTACCCCTAGAAAT 1140
Db 1081 CAGGAAAAAGAAATCTAGTTTAAATGAGGCTACAGAGGAAATGAAGGTACCCCTAGAAAT 1140
QY 1141 CCATCAATGAGAAGACTTTTTCAGCTACTATCGTCAATTTGAAATGCGCCCTTTAAAGGAC 1200
Db 1141 CCATCAATGAGAAGACTTTTTCAGCTACTATCGTCAATTTGAAATGCGCCCTTTAAAGGAC 1200
QY 1201 GCAGAGTTGAAGGAATTCATATGAAGACATTTCTAAACTTGCCTGAGTTTGAAAAATTTG 1260
Db 1201 GCAGAGTTGAAGGAATTCATATGAAGACATTTCTAAACTTGCCTGAGTTTGAAAAATTTG 1260
QY 1261 TTACAAATACATTAATCTTAAAGCATCAAAATTTGGTGTTCGCCAAGGCATTTAGAGACT 1320

Db 1261 TTAAACAATACATTAATAATCTTTAAAGCATCAAAATTTGGTGTTCGCCAAGGCATTTAGAGACT 1320
QY 1321 CTACTGTGTTAGGCTATATCTTTTGTATAAAAACAACAGGTTTTTTGAAAAATATTACTGT 1380
Db 1321 CTACTGTGTTAGGCTATATCTTTTGTATAAAAACAACAGGTTTTTTGAAAAATATTACTGT 1380
QY 1381 ATAGTGTTCAGCTTAACTTTTGAGAGAAATTTAATATGCTCATGAGGTTATCAAACTAT 1440
Db 1381 ATAGTGTTCAGCTTAACTTTTGAGAGAAATTTAATATGCTCATGAGGTTATCAAACTAT 1440
QY 1441 GTAATTTTGTCTGTTTATTTTCTTCCCTTTGTAATTTTACTTGTATGTTTATATCTTC 1500
Db 1441 GTAATTTTGTCTGTTTATTTTCTTCCCTTTGTAATTTTACTTGTATGTTTATATCTTC 1500
QY 1501 ATTAAGAATGTTATTAATAAAAAAAA 1528
Db 1501 ATTAAGAATGTTATTAATAAAAAAAA 1528
RESULT 2
HSAJ5273
LOCUS HSAJ5273 1528 bp mRNA linear PRI 13-DEC-2000
DEFINITION Homo sapiens mRNA for Kin17 protein.
ACCESSION AJ005273
VERSION AJ005273.1 GI:3850703
KEYWORDS KIN17 gene.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1
AUTHORS Kannouche, P., Mauffrey, P., Pinon-Lataillade, G., Mattei, M.G.,
Sarasin, A., Daya-Grosjean, L. and Angulo, J.F.
TITLE Molecular cloning and characterization of the human KIN17 cDNA
encoding a component of the UVC response that is conserved among
metazoans
JOURNAL Carcinogenesis 21 (9), 1701-1710 (2000)
PUBMED 10964102
REFERENCE 2
AUTHORS Mauffrey, P.
TITLE Direct Submission
JOURNAL Submitted (08-APR-1998) Mauffrey P., LGR, CEA-DSV-DRR, BP 6 -
Fontenay aux Roses Cedex, 92265, FRANCE
REMARK Revised by [3]
REFERENCE 3 (bases 1 to 1528)
AUTHORS Mauffrey, P.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1998) Mauffrey P., LGR, CEA-DSV-DRR, BP 6 -
Fontenay aux Roses Cedex, 92265, FRANCE
COMMENT On Nov 7, 1998 this sequence version replaced gi:3046739.
Kannouche P., Mauffrey P., Pinon-Lataillade G., Biard D.S.F.,
Angulo J.F.
Sequences codant pour la Proteine kin17 et leurs Applications -
French Patent Nr 97 15536 1:1-50(1997).
FEATURES
source
1. .1528
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/sex="male"
/tissue_type="testis"
1. .1528
/gene="Kin17"
66. .1247
/gene="Kin17"
/codon_start=1
/protein_id="CAA06462.1"
/db_xref="GI:3850704"
/db_xref="GOA:O60870"
/db_xref="UniProt/TREMBL:O60870"
translation="MGKSDFLTPKAIANRIKSKGLQKLWYQWQKQCRDENGFKCH"

CHSESHQQLLLASENPOFMDYSEBERNDPFLERRRFGTKRVHNNIVVYEXISHR
 EHHNNAQTWETLIDFTKWLREGGLCKVDETPKWHYQYIDRDPETIRRLQLEKKK
 QLDDBEKTAFTKIFGSSRGLGKEQEVPTFELSRENDSEKVFNFNLSSSGAT
 SSKSTLGSFSAKITGSSASVKRESQSSTOSKSKKKXKXKXKXKXKXKXKXKX
 TDYWLQPEIIVKIIITKLGKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX
 VIPAPGKRILVNGSYRGNEGTLSINEKTFSATIVETGPKLGRVRVGIQVEDISK
 A

ORIGIN

Query Match	100.0%	Score 1528	DB 8	Length 1528
Best Local Similarity	100.0%	Pred. No. 2e-311		
Matches 1528	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	CTAGAATTTCAGGCGCGCTGAATTCAGAACTGGGGTCCAGAAAGTGATCGCTGGCGTGG	60	
Db	1	CTAGAATTTCAGGCGCGCTGAATTCAGAACTGGGGTCCAGAAAGTGATCGCTGGCGTGG	60	
QY	61	TCGCCATGGGGAAGTCGGAATTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCA	120	
Db	61	TCGCCATGGGGAAGTCGGAATTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCA	120	
QY	121	AGGGGCTGCAGAAAGCTACGCTGGTATTCGCAGATGTGCAGAGCAGTGCCGGGACGAGA	180	
Db	121	AGGGGCTGCAGAAAGCTACGCTGGTATTCGCAGATGTGCAGAGCAGTGCCGGGACGAGA	180	
QY	181	ATGGCTTTAAGTGTCAATTCGATGTCGGAATCTCATCAGAGACAATATTCGCTGGCTTCAG	240	
Db	181	ATGGCTTTAAGTGTCAATTCGATGTCGGAATCTCATCAGAGACAATATTCGCTGGCTTCAG	240	
QY	241	AAATCTCAGCAGATTTATGGATATTTTTCAGAGGAATTCGGAATGACTTTCTAGAAC	300	
Db	241	AAATCTCAGCAGATTTATGGATATTTTTCAGAGGAATTCGGAATGACTTTCTAGAAC	300	
QY	301	TTCTCAGGAGAGCGCTTTGGCAGCTTAAAGGGTCCACAAACATGTCGTACAACGAATACA	360	
Db	301	TTCTCAGGAGAGCGCTTTGGCAGCTTAAAGGGTCCACAAACATGTCGTACAACGAATACA	360	
QY	361	TCAGCCACCGAGAGCAGTCATCCATGAATGCCATCTCAGTGGGAACTCTGACTGATTTTA	420	
Db	361	TCAGCCACCGAGAGCAGTCATCCATGAATGCCATCTCAGTGGGAACTCTGACTGATTTTA	420	
QY	421	CTAAGTGGCTGGGACAGAGAGGCTTCGCAAGTGGACGAGACACCAAAAGGCTGGTATA	480	
Db	421	CTAAGTGGCTGGGACAGAGAGGCTTCGCAAGTGGACGAGACACCAAAAGGCTGGTATA	480	
QY	481	TTCAAGTACATAGACAGGAGCCCAAGAACTATCCGCGGCACTTGGAACTGGAGAAAAAGA	540	
Db	481	TTCAAGTACATAGACAGGAGCCCAAGAACTATCCGCGGCACTTGGAACTGGAGAAAAAGA	540	
QY	541	AAAAGCAGGACCTTGATGATGAAGAAAAAACTGCCAAATTTATTTGAAGAGCAAGTGAGAA	600	
Db	541	AAAAGCAGGACCTTGATGATGAAGAAAAAACTGCCAAATTTATTTGAAGAGCAAGTGAGAA	600	
QY	601	GAGGCTGGAGGAGGAGGAGGCTCCCTACTTTTACGGAAATTAAGCAGAGAAAAATG	660	
Db	601	GAGGCTGGAGGAGGAGGAGGCTCCCTACTTTTACGGAAATTAAGCAGAGAAAAATG	660	
QY	661	ATGAAGAGAAAGTCAGCTTTAAATTTTGAAGTAAGGAGCATGTAGCTCATCCGGAGCAACAT	720	
Db	661	ATGAAGAGAAAGTCAGCTTTAAATTTTGAAGTAAGGAGCATGTAGCTCATCCGGAGCAACAT	720	
QY	721	CTTCCAAAGTCAAGTACTCTGGGACCGAGTGCATCTGAAGAGCATAGGAGTTTCAGCATCAG	780	
Db	721	CTTCCAAAGTCAAGTACTCTGGGACCGAGTGCATCTGAAGAGCATAGGAGTTTCAGCATCAG	780	
QY	781	TGAACGAAAAGATCTTCCAGAGCTCACTCAGTCTAAGAAAAAGAAAAAGAAAT	840	
Db	781	TGAACGAAAAGATCTTCCAGAGCTCACTCAGTCTAAGAAAAAGAAAAAGAAAT	840	
QY	841	CTGCACTGGATGAATCATCGAGATTGAAGAGGAAAAAGAAAGAACTGCCCGAACAGACT	900	
Db	841	CTGCACTGGATGAATCATCGAGATTGAAGAGGAAAAAGAAAGAACTGCCCGAACAGACT	900	

QY	901	ACTGGCTACAGCCTGAAATTTTGTGAAAAATATAACAAGAACTGGGAGAGAAATATC	960
Db	901	ACTGGCTACAGCCTGAAATTTTGTGAAAAATATAACAAGAACTGGGAGAGAAATATC	960
QY	961	ATAAGAAAAAGGCTATTGTTAAGGAAGTAATTGACAAATATACAGCTGTTGTAAGATGA	1020
Db	961	ATAAGAAAAAGGCTATTGTTAAGGAAGTAATTGACAAATATACAGCTGTTGTAAGATGA	1020
QY	1021	TTGATTTCTGGAGAGCAAGCTGAAACTTTGACAGACATCTTTTAGAGACAGTAATTTCCAGCAC	1080
Db	1021	TTGATTTCTGGAGAGCAAGCTGAAACTTTGACAGACATCTTTTAGAGACAGTAATTTCCAGCAC	1080
QY	1081	CAGAAAAAAGAAATCTTACTGTTTTAAATGGAGGCTACAGAGGAAATGAAGTACCCCTAGAAAT	1140
Db	1081	CAGAAAAAAGAAATCTTACTGTTTTAAATGGAGGCTACAGAGGAAATGAAGTACCCCTAGAAAT	1140
QY	1141	CAATCAATGAGAGACTTTTTTCAGCTACTATCGTCAATTTGAAACTGGCCCTTTTAAAGGAC	1200
Db	1141	CAATCAATGAGAGACTTTTTTCAGCTACTATCGTCAATTTGAAACTGGCCCTTTTAAAGGAC	1200
QY	1201	GCAGAGTTGAAAGGAATTCATATATGAAGACATTTCTTAAACTTGCCTGAGTTTGAAATTTTG	1260
Db	1201	GCAGAGTTGAAAGGAATTCATATATGAAGACATTTCTTAAACTTGCCTGAGTTTGAAATTTTG	1260
QY	1261	TTAAACAATACATTAATAATCTTTAAAGCATCAAAATTCGCTGTTGCCAAGGCATTTAGACT	1320
Db	1261	TTAAACAATACATTAATAATCTTTAAAGCATCAAAATTCGCTGTTGCCAAGGCATTTAGACT	1320
QY	1321	CTACTGTGTTAGGGTATATCTTTTGTATTAACAAACAGGTTTTTGAATAATTTACTGT	1380
Db	1321	CTACTGTGTTAGGGTATATCTTTTGTATTAACAAACAGGTTTTTGAATAATTTACTGT	1380
QY	1381	ATAGTGTGTCAGTAAACTTTTGTATTAACAAACAGGTTTTTGAATAATTTACTGT	1440
Db	1381	ATAGTGTGTCAGTAAACTTTTGTATTAACAAACAGGTTTTTGAATAATTTACTGT	1440
QY	1441	GTAAATTTGCTCTGTTATTTTGTGTTTCTTCTGTTTAAATTTTACTGTAGTGTATATCTTC	1500
Db	1441	GTAAATTTGCTCTGTTATTTTGTGTTTCTTCTGTTTAAATTTTACTGTAGTGTATATCTTC	1500
QY	1501	ATTAAGAAATGTTATTTATTAATAAAAAA	1528
Db	1501	ATTAAGAAATGTTATTTATTAATAAAAAA	1528

RESULT 3

BC017309	Homo sapiens KIN, antigenic determinant of recA protein homolog (mouse), mRNA (cdna clone MGC:29595 IMAGE:5089578), complete cds.	2069 bp	mRNA	linear	PRI 29-JUN-2004
LOCUS	BC017309	2	GI:40226034		
DEFINITION	BC017309				
ACCESSION	BC017309				
VERSION	BC017309.2				
KEYWORDS	MGC.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutharia; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 2069)				
AUTHORS	Klausner, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Schuler, G.D., Altshuler, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Schetz, T.E., Brownstein, M.J., Usdin, T.B., Loquellano, N.A., Peters, G.J., Carninci, P., Prange, C., Raha, S.S., Bosak, S.A., McEwan, P.J., Abramson, R.D., Mullaby, S.J., Bosak, S.A., Gunaratne, P.H., Richards, S., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Hulyk, S.W., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Lu, X., Gibbs, R.A., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,				

Bouffard G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalhus, D.E.,
Schnierch, A., Schein, J.E., Jones, S.J., and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932
2 (bases 1 to 2069)

Strauberg, R.

Direct Submission

Submitted (05-NOV-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

On Dec 19, 2003 this sequence version replaced gi:16878214.

Contact: MGC help desk

Email: cgaps-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC),

Gaithersburg, Maryland;

Web site: <http://www.nisc.nih.gov/>

Contact: nisc.mgc@nih.gov

Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,

Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,

Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,

Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,

Maduro, Q.B., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,

McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,

Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,

Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IMAGE Plate: 39 Row: 0 Column: 18

This clone was selected for full length sequencing because it

passed the following selection criteria: matched mRNA gi: 13124882.

FEATURES

source

1..2069
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:29595 IMAGE:5089578"
/tissue_type="Pancreas, epithelioid carcinoma"
/clone_lib="NIH MGC_42"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
1..2069
/gene="KIN"
/note="synonyms: KIN17, BTCD"
/db_xref="GeneID:22944"
/db_xref="MIM:601720"
12..1193

gene

/codon_start=1
/product="HsKIN17 protein"
/protein_id="AAH17309.1"
/db_xref="GI:16878215"
/db_xref="GeneID:22944"
/db_xref="MIM:601720"

CDS

/translation="MGKSDFLTPKAIANRIKSGIKQLRWYCOMQKCRDENGFKCH
CMSSSHQQLLASNPQFMDYFSEPRNDFLELRFRFTGKRVHNNVYVEYISHR
EHIHNAQTQWLTDFDKWGLRGCLKVDETPKGWIQYIDRDETIRQOLEKXKX
QDLDEETAKFIEQVNRGLGKEQVPTFTLSRENDEKVFNLKSGACSSGAT
SSKSTLTGPSALKTIGSSASVKRKESSQTSQSKKKKXGALDEIMEIEBKRTAR
TDYWLQPEIIVKIIITKKLGEYHKKAIKVEIDKYTAVRWMDSGDKLKDQTHLET
VIPAPGKRILLVNGYRGNEGTLESINEKTFESATIVETGPLKGRRVEGIQYEDISKL
A"

ORIGIN

Query Match	95.1%	Score 1453;	DB 8;	Length 2069;
Best Local Similarity	95.7%	Pred. No. 1.2e-295;		
Matches 1467;	Conservative 0;	Mismatches 0;	Indels 4;	Gaps 1;
QY	55	CCGTGGTCGCCATGGGGAAGTCGGATTTCTTACTCCCAAGGCTATCGCCAAACAGGATCA	114	
Db	1	CCGTGGTCGCCATGGGGAAGTCGGATTTCTTACTCCCAAGGCTATCGCCAAACAGGATCA	60	
QY	115	AGTCCAAAGGGGCTGCAGAAGCTACGCTGGTATTTGCCAGATGTGCCAGAAGCAGTCCGGG	174	
Db	61	AGTCCAAAGGGGCTGCAGAAGCTACGCTGGTATTTGCCAGATGTGCCAGAAGCAGTCCGGG	120	
QY	175	ACGAGAATGGCTTTAAGTGTCAATGTATGCCGATCTCATCAGAGACAATATTGCTGG	234	
Db	121	ACGAGAATGGCTTTAAGTGTCAATGTATGCCGATCTCATCAGAGACAATATTGCTGG	180	
QY	235	CTTCAGAAAATCCTCAGCAGCTTTATGGAATTTTTCAGAGGAATTCGGAATGACTTTC	294	
Db	181	CTTCAGAAAATCCTCAGCAGCTTTATGGAATTTTTCAGAGGAATTCGGAATGACTTTC	240	
QY	295	TAGAACTTCTCAGGAGACGCTTTGGCACTAAAAGGGTCCACAACAACATTTCTACAACG	354	
Db	241	TAGAACTTCTCAGGAGACGCTTTGGCACTAAAAGGGTCCACAACAACATTTCTACAACG	300	
QY	355	AATACATCGCCACCGAGAGACATCCACATGAATGCCATCAGTGGGAAATCTGACTG	414	
Db	301	AATACATCGCCACCGAGAGACATCCACATGAATGCCATCAGTGGGAAATCTGACTG	360	
QY	415	ATTTTACTAGTGGCTCGGAGAGAGGCTTTGTCAAAGTGGAGAGACACCAAAAGGCT	474	
Db	361	ATTTTACTAGTGGCTCGGAGAGAGGCTTTGTCAAAGTGGAGAGACACCAAAAGGCT	420	
QY	475	GGTATATTTCAGTACATAGACAGGACCCAGAAATCATCCCGCGCAATTCGAACTGGAGA	534	
Db	421	GGTATATTTCAGTACATAGACAGGACCCAGAAATCATCCCGCGCAATTCGAACTGGAGA	480	
QY	535	AAAAGAAAAGCAGGACCTTGATGATGAAGAAAATTCGCAATTTATTGAAGAGCAAG	594	
Db	481	AAAAGAAAAGCAGGACCTTGATGATGAAGAAAATTCGCAATTTATTGAAGAGCAAG	540	
QY	595	TGAGAAGAGCCTCGAAGGGAAGGAACAGAGGTCCTACTTTTACCGAATTAAGCAGAG	654	
Db	541	TGAGAAGAGCCTCGAAGGGAAGGAACAGAGGTCCTACTTTTACCGAATTAAGCAGAG	600	
QY	655	AAAATGATGAAGAGAAAGTCACGTTTAAATTTGAGTAAAGGAGCATGTAGCTCATCCGGAG	714	
Db	601	AAAATGATGAAGAGAAAGTCACGTTTAAATTTGAGTAAAGGAGCATGTAGCTCATCCGGAG	660	
QY	715	CAACATCTTCCAGTCAAGTACTCTGGGACCGAGTGCATGAGACCATAGGAAGTTTCAG	774	
Db	661	CAACATCTTCCAGTCAAGTACTCTGGGACCGAGTGCATGAGACCATAGGAAGTTTCAG	720	
QY	775	CATCAGTGAAGCAAGAAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAAGAAAAGAGAAA	834	
Db	721	CATCAGTGAAGCAAGAAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAAGAAAAGAGAAA	780	
QY	835	AGAAATCTGCATCGGATGAAATCATGGAGATTGAAGAGGAAAGAAAAGAAAGAACTCCCGAA	894	
Db	781	AGAAATCTGCATCGGATGAAATCATGGAGATTGAAGAGGAAAGAAAAGAAAGAACTCCCGAA	840	
QY	895	CAGACTACTGGCTACAGCTGAAATTTATTTGAAATTTATAACCAAGAACTGGAGAGA	954	
Db	841	CAGACTACTGGCTACAGCTGAAATTTATTTGAAATTTATAACCAAGAACTGGAGAGA	900	
QY	955	AATATCAAGAAAAGGCTATTGTTAAGGAAGTAAATTTGACAAATATACAGCTGTGTGA	1014	
Db	901	AATATCAAGAAAAGGCTATTGTTAAGGAAGTAAATTTGACAAATATACAGCTGTGTGA	960	
QY	1015	AGATGATTGATTCTGGAGACAAGCTGAAACTTGACCAAGCTCACTTTAGAGACAGTAATTC	1074	
Db	961	AGATGATTGATTCTGGAGACAAGCTGAAACTTGACCAAGCTCACTTTAGAGACAGTAATTC	1020	

```

QY 1075 CAGCACCAGGAAAGAAATCTAGTTTTAAATGGAGGCTACAGAGAAATGAAGTACCC 1134
DB 1021 CAGCACCAGGAAAGAAATCTAGTTTTAAATGGAGGCTACAGAGAAATGAAGTACCC 1080
QY 1135 TAGAATCCATCAATGAGAAGACTTTTTAGCTACTATCGTCAATGAAACTGGCCCTTTAA 1194
DB 1081 TAGAATCCATCAATGAGAAGACTTTTTAGCTACTATCGTCAATGAAACTGGCCCTTTAA 1140
QY 1195 AAGGACGAGAGTGAAGGAATCAATATGAAGACATTTCTAAACTGGCTGAGTTGAA 1254
DB 1141 AAGGACGAGAGTGAAGGAATCAATATGAAGACATTTCTAAACTGGCTGAGTTGAA 1200
QY 1255 AATTTGTTAAACAATACATTTAAATCTTAAAGCATCAAAATTTGTTTCGCAAGGCATTAT 1314
DB 1201 AATTTGTTAAACAATACATTTAAATCTTAAAGCATCAAAATTTGTTTCGCAAGGCATTAT 1260
QY 1315 GAGACTCTACTGTTAGGAGTATTTCTTTTGTATATAAAACAAACAGGTTTTTGAATAAT 1374
DB 1261 GAGACTCTACTGTTAGGAGTATTTCTTTTGTATATAAAACAAACAGGTTTTTGAATAAT 1320
QY 1375 TACTGTA-----TAGTTGTTTACGTAACCTTTGAGAAGAAATTTAATATGTCATGAGT 1430
DB 1321 TACTGTAATAGTTAGTTGTTTACGTAACCTTTGAGAAGAAATTTAATATGTCATGAGT 1380
QY 1431 ATCAAACTATGTAATTTGCTGCTGTTATTTTGTGTTTCTTTGTAATTTTACTTTGATGAT 1490
DB 1381 ATCAAACTATGTAATTTGCTGCTGTTATTTTGTGTTTCTTTGTAATTTTACTTTGATGAT 1440
QY 1491 TTATATCTTTCAATTAAGAAATGTTATTATATAA 1521
DB 1441 TTATATCTTTCAATTAAGAAATGTTATTATATAA 1471

RESULT 4
AX003311
LOCUS AX003311 1002 bp DNA linear PAT 24-AUG-2000
DEFINITION Sequence 4 from Patent WO9929845.
ACCESSION AX003311
VERSION AX003311.1 GI:9927128
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE
1 Angulo-Mora, J.F. and Mauffrey, P.
AUTHORS Sequences coding for kin17 protein and their applications
TITLE Patent: WO 9929845-A 4 17-JUN-1999;
JOURNAL ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE
(FR)
FEATURES
source
1. .1002
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 65.6%; Score 1002; DB 6; Length 1002;
Best Local Similarity 100.0%; Pred. No. 1.4e-200;
Matches 1002; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 215 TCAGAGACAACATATGCTGCTTCAGAAAATCTCAGCAGTATGATATTTTTCAGA 274
DB 1 TCAGAGACAACATATGCTGCTTCAGAAAATCTCAGCAGTATGATATTTTTCAGA 60
QY 275 GGAATCCGAATGACATTTCTAGAACTCTCAGAGAGCGTTTGGCACTAAAGGTTCCA 334
DB 61 GGAATCCGAATGACATTTCTAGAACTCTCAGAGAGCGTTTGGCACTAAAGGTTCCA 120
QY 335 CAACAACATTTGCTACAAACAAATACATCAGCCACCGAGAGACATCCACATGAATGCCAC 394
DB 121 CAACAACATTTGCTACAAACAAATACATCAGCCACCGAGAGACATCCACATGAATGCCAC 180

```

```

QY 395 TCAGTGGGAAACTCTGACTGATTTTACTAAGTGGCTGGGACAGAGAGCTTTGTGCAAAAGT 454
DB 181 TCAGTGGGAAACTCTGACTGATTTTACTAAGTGGCTGGGACAGAGAGCTTTGTGCAAAAGT 240
QY 455 GGAAGAGACACAAAGGCTGGTATATTTAGTATACATAGACAGGAGCCAGAGAACTATCCG 514
DB 241 GGAAGAGACACAAAGGCTGGTATATTTAGTATACATAGACAGGAGCCAGAGAACTATCCG 300
QY 515 CGGCAACTGGAACCTGGAGAGAAAAGAAAAGCAGGACCTTTGATGATGAAGAAAACATGCG 574
DB 301 CGGCAACTGGAACCTGGAGAGAAAAGAAAAGCAGGACCTTTGATGATGAAGAAAACATGCG 360
QY 575 CAATTTATTTGAAGAGCAAGTGAAGAGGCTTGAAGGGAAGGAAACAGGAGTCCCTAC 634
DB 361 CAATTTATTTGAAGAGCAAGTGAAGAGGCTTGAAGGGAAGGAAACAGGAGTCCCTAC 420
QY 635 TTTTACGGAATTAAGCAGAGAAAATGATGAAGAGAAAGTCAAGTTTAAATTTGAGTAAAGG 694
DB 421 TTTTACGGAATTAAGCAGAGAAAATGATGAAGAGAAAGTCAAGTTTAAATTTGAGTAAAGG 480
QY 695 AGCATGTAGCTCATCCGGAGCAACATCTTCCAAAGTCAAGTACTCTGGGACCGAGTGCCT 754
DB 481 AGCATGTAGCTCATCCGGAGCAACATCTTCCAAAGTCAAGTACTCTGGGACCGAGTGCCT 540
QY 755 GAAGACGATAGGAAGTTCAGCATCAGTGAACCAAGAAATCTTCCAGAGCTCAACTCA 814
DB 541 GAAGACGATAGGAAGTTCAGCATCAGTGAACCAAGAAAGTCTTCCAGAGCTCAACTCA 600
QY 815 GTCTAAAGAAAAGAAAAGAAAATCTGCACTGATGAATCATGAGAGATTTGAAGAGGA 874
DB 601 GTCTAAAGAAAAGAAAAGAAAATCTGCACTGATGAATCATGAGAGATTTGAAGAGGA 660
QY 875 AAAGAAAAGAACTGCGGACAGACTACTGCGCTCAGAGCTGAAATTTATTTGTAATAATAT 934
DB 661 AAAGAAAAGAACTGCGGACAGACTACTGCGCTCAGAGCTGAAATTTATTTGTAATAATAT 720
QY 935 AACCAAGAACTGGGAGAGAAAATATCATAAAGAAAAGGCTATTGTTAAGGAAGTAAATGA 994
DB 721 AACCAAGAACTGGGAGAGAAAATATCATAAAGAAAAGGCTATTGTTAAGGAAGTAAATGA 780
QY 995 CAAATATACAGCTGTTGGAAGATGATGATTTCTGAGACAAAGCTGAAACTTGACACAGAC 1054
DB 781 CAAATATACAGCTGTTGGAAGATGATGATTTCTGAGACAAAGCTGAAACTTGACACAGAC 840
QY 1055 TCATTTAGACAGACTAATTTCCAGCACCAGGAAAAGAAATTTCTAGTTTAAATGGAGGCTA 1114
DB 841 TCATTTAGACAGACTAATTTCCAGCACCAGGAAAAGAAATTTCTAGTTTAAATGGAGGCTA 900
QY 1115 CAGAGAAAATGAAGGTACCTAGATCCATCAATGAGAGAGCTTTTTCAGCTACTATCGT 1174
DB 901 CAGAGAAAATGAAGGTACCTAGATCCATCAATGAGAGAGCTTTTTCAGCTACTATCGT 960
QY 1175 CATTGAAACTGGCCCTTTTAAAGGACGAGAGTGAAGGAAT 1216
DB 961 CATTGAAACTGGCCCTTTTAAAGGACGAGAGTGAAGGAAT 1002

RESULT 5
MMKIN17
LOCUS Mouse kin17 mRNA for kin17 protein.
DEFINITION X58472
ACCESSION X58472.1 GI:1045209
VERSION kin17 gene; kin17 protein; zinc-finger protein.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 1414)
AUTHORS Angulo,J.F., Rouer,E., Benarous,R. and Devoret,R.
TITLE Identification of a mouse cDNA fragment whose expressed polypeptide

```


[illegible]

Db	1377	AAAAAAAA	1384
RESULT 7			
AX003331			
LOCUS	AX003331	1390 bp	DNA linear PAT 24-AUG-2000
DEFINITION	Sequence 24 from Patent WO929845.		
ACCESSION	AX003331		
VERSION	AX003331.1	GI:9927146	
KEYWORDS			
SOURCE	Mus sp.		
ORGANISM	Mus sp.		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;		
	Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1		
AUTHORS	Angulo-Mora, J.F. and Mauffrey, P.		
TITLE	Sequences coding for kin17 protein and their applications		
JOURNAL	Patent: WO 9929845-A 24 17-JUN-1999;		
	ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE		
(FR)			
FEATURES	Location/Qualifiers		
source	1..1390		
	/organism="Mus sp."		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:10095"		
ORIGIN			
	Query Match 64.8%; Score 990.6; DB 6; Length 1390;		
	Best Local Similarity 83.0%; Pred. No. 3.5e-198;		
	Matches 1144; Conservative 0; Mismatches 229; Indels 6; Gaps 1;		
Qy	66	ATGGGGAAGTCGCAGATTCTTACTCCCAAGCTATCGCCAAGGATCAAAGTCCAAGGGG	125
Db	1	ATGGCNAATCGAATTTCTGAGCCCCAGGCCCATCGCCAATAGAATAAGTCCAAAGG	60
Qy	126	CTGCAGAAGCTACGCTGGTAATGCCAGATGTGCCAAGACGAGTGCCGGGACGAGAATGCG	185
Db	61	CTCCAGAAGCTTCGCTGGTAGTCCAGATGTGCCAAAAGCAATGCCGACGAGAATGCG	120
Qy	186	TTTTAAGTGTCATTTGTATGCCGAATCTCATCAGAGCAACTATTGTCTGGCTTCAGAAAT	245
Db	121	TTTAAAGTGTCACATGTATGTCGAATCTCATCAAGACAACCTGTGTGCTTCAGAAAA	180
Qy	246	CCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTCGGAATGACTTTCTAGAACCTTCTC	305
Db	181	CCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTCGGAATGACTTTCTGGAACTCTG	240
Qy	306	AGGAGACGCTTTGGCACTAAAGGGTCCACAAACAATTTCTACACGAATACATCAGC	365
Db	241	AGGCGACGCTTTGGCACTAAAGGGTCCACAAACAATTTCTACAAATGAATACATCAGC	300
Qy	366	CACCGAGACACATCCACATGAATGCCACTCAGTGGGAACTCTGACTGATTTTACTAAG	425
Db	301	CACCGAGACACATCCACATGAACGCTACCAGTGGGAGACACTGACCGACTTTACCAAG	360
Qy	426	TGCGTGGCAGAGAGGCTTTGTCCAAAGTGGACGAGACACCAAAGGCTGGTATATTTCAG	485
Db	361	TGGCTGGCAGAGAGGCTTTGTGAAGTGGATGAGCACCCGAAGGCTGGTACTATTTCAG	420
Qy	486	TACATGACAGGACCCAGAAAATCTCCGCGGCAACTGGAACTGGAGAAAAAGAAAAAG	545
Db	421	TACATGACAGAGACCCAGAAAACCATCCGTCGCGCAACTGGAAATTAGAAAAAGAGAAG	480
Qy	546	CAGACCTTGATGATGAAGAAAAAATGCCCCAAATTTATTGAAGACGCAAGTGAGNAGGC	605
Db	481	CAAGATCTGACGATGAAGAAAAAATCTGCCAAGTTCAATTGAGGACGAGGTGAGNAGGC	540
Qy	606	CTGMAAGGGAAGGACAGGAGGTCCTTACTTTTACGGAAATTTAAGCAGAGAAAAATGATGA	665
Db	541	CTGMAAGGGAAGGACAGGAGACACCTTGTTTTTACAGAACTTAGCCGAGAAATGAGAA	600
Qy	666	GAGAAAGTCAGTTTTTAATTTTGAAGTAAAGGAGCATGTAGCTCATCCGAGCAACATCTTCC	725


```

Db      601  GAAAGTACGTTCAATCTGATTAAGAGCGGGTGGCTCAGCGGAGCTACAACATCC 660
QY      726  AAGTCAAGTACTCTGGGACCGAGTGCATGAAGACGATAGGAAGTTTCAGCATCAGTGAAA 785
Db      661  AAGTCAAGCTCTTTGGGACCAAGTGCATCTGAAGCTGCTGGGGAGCGCAGCATCCGGGAAA 720
QY      786  CGAAGAGATCTTCCAGAGCTCAACTCAGTCTAAAGAAAGAAAGAAAGAAATCTGCA 845
Db      721  CGAAGAGATCTTCAAGAGCTCCGCCAGCTGC-----GAAGAGAGAGATCGGCC 774
QY      846  CTGGATGAATCATGTGAGATTGAAGAGGAAAAAGAAAGAACTGCCCGAACAGACTACTGG 905
Db      775  CTGGATGAGATCATGTGAGCTCGAAGAGGAAAAAGAAAGAAAGCAAGCGGACAGCGCTGG 834
QY      906  CTACAGCTGAAATTTATTTGTGAAAAATTATACCAAGAAACTGGGAGAGAAATATCATAG 965
Db      835  TTACAGCCGGGATCGTTGTGAAAAATTATACGAAGAAGCTTTGGGAGAAATATCACAG 894
QY      966  AAAAAAGCTATTGTTAAGGAAGTAATTGACAAATATACAGCTGTTGTGAAGATGATTGAT 1025
Db      895  AAGAAAGGGTTCGTTAAGGAAGTATTGACAGTACACAGTGTGTTAAGATGACTGAC 954
QY      1026  TCTGGAGACAAGCTGAAACTTGACCAAGACTCATTTAGAGACAGTAAATCCAGCACCAGGA 1085
Db      955  TCTGGAGACAAGCTGAAACTTGACCAAGACTCATTTAGAGACAGTCAATTCGGCCCGGGG 1014
QY      1086  AAAAGAAATCTAGTTTAAATGAGGCTACAGAGAAATGAAGGTACCTAGATTCATC 1145
Db      1015  AAAAGGGTCTAGTTTAAATGAGGCTACAGAGAAATGAAGGCACTCTCGAATCCATC 1074
QY      1146  AATGAGAAGACTTTTTCAGCTACTATCTGATGAACTGAACTGGCCCTTTAAAGAGACGCA 1205
Db      1075  AATGAGAAGCTTTTTCAGCCAGTAGTATGAACTGAACTTTGAAGAGACGCGCA 1134
QY      1206  GTTGAAGGAATTCATATGAAGACATTTCTAAACTTGCCTGAGTTTGAAGAAATTTGTTAAC 1265
Db      1135  GTTGAAGGTATTCATATGAGACATATCTAACTTGTCTGAGTTTGAAGAAATTTGATAC 1194
QY      1266  AATACATTAATAATCTTAAAGCATCAAAATGGTGTTCGCCAAGCATATGAGACTTACT 1325
Db      1195  AACACATTTGAACCTGTGAAGCATCAAAATGGTGTTCGCCAAGCACTGTGTAACCTACT 1254
QY      1326  GTGTTAGGTATATCTTTTGTATATAAACAACAGGTTTGTGAATATATTACTGTATAGT 1385
Db      1255  GTGTTAGGGGATTTGTTGTATTAATAAAGAAAAAATCATCTATTATAACTAGTGAA 1314
QY      1386  TGTTCAGCTAAACTTTGAGAAGAAATTAATATGCTCATGAGGTATCAAACTATGTAA 1444
Db      1315  TAGTGGGTAAATTTATATAAATCTATGTTTTTTTAAAGTGAAGAAAAA 1373

RESULT 8
AC119716
LOCUS   Rattus norvegicus clone CH230-274116, *** SEQUENCING IN PROGRESS
DEFINITION
AC119716
VERSION AC119716.6 Gi:25092366
KEYWORDS HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
SOURCE  Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidae; Muridae; Murinae; Rattus.
1 (bases 1 to 221547)
Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooke,S., Amin,A., Angiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Blawalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,

```

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 JOURNAL

Unpublished
 2 (bases 1 to 221547)
 Worley,K.C.
 Direct Submission

Submitted (01-MAY-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

REFERENCE
 AUTHORS
 TITLE
 JOURNAL

Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Nov 19, 2002 this sequence version replaced gi:23611286.
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>

Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
 Davila,M.L., Davis,C., Davy-Carroll,L., De Andra,C., Dederich,D.,
 Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
 Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Evans,K.,
 Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
 Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
 Gregorogis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
 Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
 Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
 Hernandez,R., Hines,S., Hladun,S.B., Hodgson,A., Hogue,M.,
 Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
 Karkathy,S., Kelly,S., Khan,Z., King,L., King,L., Kovar,C.,
 Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
 Lorenshuwa,L., Louissegh,H., Lozado,R.J., Lu,X., Ma,J.,
 Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
 Mangum,B., Mapa,P., Martin,K., Martin,R., Martinez,E.,
 Mahoney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
 Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
 Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
 Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
 Nwackelmech,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
 Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
 Platter,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,-L.,
 Puato,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
 Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
 Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
 Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Tjorjoe,Z., Umanai,K.,
 Valas,R., Vera,V., Villalana,D., Waldron,L., Walker,B., Wang,J.,
 Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
 Williams,G., Willison,R., Wleczkyk,R., Wooden,H., Worley,K.,
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
 Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
 Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
 Weinstein,G. and Gibbs,R.A.
 Direct Submission

clone_end:77 vector_side:right		49.7%; Score 759; DB 14; Length 130262; 77.1%; Pred. No. 1.2e-149; Matches 1060; Conservative 0; Mismatches 260; Indels 55; Gaps 9;	
ORIGIN	Query Match		
	Best Local Similarity		
	Matches 1060; Conservative 0; Mismatches 260; Indels 55; Gaps 9;		
QY	46 TGATCGCTGCCGTGCTGCCATGCGGAAGTGGATTTTCTTACTCCCAAGGCTATCGCCA 105		
DB	99745 TGGTCGGGCCAGGCTAGCTTGTGCAAGGGCG-TTTTTCTTAGCCCTAAGGCCATCGCA 99803		
QY	106 ACAGGA-TCAAGTCCAAAGGGCTGAGAGCTACGCTGTATGTCAGATGTCGCAGAG 164		
DB	99804 ATAGAAGTAAAGTCCAAAGGGCTCCAGAAAGCTTACCAGGCTGTCAGATGTCGC 99863		
QY	165 CAGTCCGGGAGCGAAATGGCTTTTAAAGTGTCTATGTCGGAATCTCATCAGAGACAA 224		
DB	99864 CAATCGCGCCACAGAGATGTCTTTAAGTGTCTATGTCGGAATCTCATCAGAGACCA 99923		
QY	225 CTATTCTGGCTTCAGAAATCCTCAGCAGTTTATGGAATTTTTCAGAGAAATCCGA 284		
DB	99924 CAGTTGCTGGCTTCAGAAAGCTCAGCAGTTTATACATTTATTTTCAGA----- 99973		
QY	285 AATGACTTCTAGAATCTTCAGGAGACGCTTTGGCACTTAAAGGCTCCACAACAACATT 344		
DB	99974 -----GCACATAAAGGGGCCCAACAACATT 99999		
QY	345 GTCTACAGCAATACATCAGCCACCGAGAGACATCCACATCAATGCCACTCAGTGGGAA 404		
DB	100000 GTCTACATGAATATATACGCCACCAAGAGACATCCACATGAATGCTACCCAGTGGGA 100059		
QY	405 ACTCTGACTGATTTTACTAAGTGGCTGGCGACAGAGAGGCTGTGCAAAAGTGACAGACA 464		
DB	100060 AAGCTGACCAACTTTTCCAAGTGGCTGGGCAGA-AGGGCTTGTGTAAGTGGGTGAGACA 100118		
QY	465 CCAAAAGGCTGTATTTTCAAGTGTACATAGACAGGACCCAGAACTATCCGCGGCAACTG 524		
DB	100119 CCAAAAGGCTGTATTTTCAAGTGTACATAGACAGGACCCAGAACTATCTGTGGGCAACTG 100177		
QY	525 GAACT-GGAGAAAAAGAAAAAGCAGACCTTGATGATGAAGAAAAAATGCCAAATTTAT 583		
DB	100178 AATATTAGAAAAAGAAAGATGGAAGATCTTGACGATGAAGAAAAAATGCCAAATTTAT 100237		
QY	584 TGAAGAGCAAGTGAAGAGGCTGGAAGGGAAGGAAAGGAGAGGCTCCCT-ACTTTACGG 642		
DB	100238 TGAGGAGCAGGTGAGAAGAGGCTGGAAGGGAAGAGAGAGAGACACCTGTTTTTACAG 100297		
QY	643 AATTAAGCAGAGAAATGATGAAGAAAGTACGTTTAAATTTGAGTAAAGGAGCATGTA 702		
DB	100298 AATTAGGACAGAAAAATTAAGGAAGAAAAATTTATGTCATCTGAATAAGGAGTGGGTG 100357		
QY	703 GCTCATCCGAGCAACATCTTCCAAGTCAAGTACTCTGGGACCGAGTGCACCTGAAGACGA 762		
DB	100358 GCTCAGCAGAGCTACAAAGTCCAGTCAAGTCTTTTGGACCAAGTGCACCTGAAGCTGC 100417		
QY	763 TAGGAAGTTTCAGCATCAGTGAACGAAAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAG 822		
DB	100418 TGAGGAGCAGCATCCGGGAATGGAAGGGTCTTTCAGAGAGCTCTCCAGCCT---G 100474		
QY	823 AAAGAAGAAAAAGAAATCTGCACTGGATCAATCATGGAGTTGAGAGGAAAAAGAAAA 882		
DB	100475 CAAAGAAGAAAGATGTCTGCCCTGGATGAATAATTATAGAGATTAAGAAGAAAAAGAAAA 100534		
QY	883 GAATGCCGCAACAGACTCTGCTACAGCTCAAGCTTAAATTTATGTAATAATTAACCAAGA 942		
DB	100535 GGACCAACAGCAGACTCTGCTTACAGCCGGGATCATTTGTGAACCTATTAAGGAGAGA 100594		
QY	943 AACTGGGAGAAATATCATAGAAGAAAGGCTATTTGTTAAGGAAGTAATTGACAAATATA 1002		
DB	100595 AGCTTGGGAGATATTCACAAGAAAAAGGGGTCAATACCGAAGAGATTGACATGTACA 100654		
QY	1003 CAGCTGTTGTAAGATGATTGATTTGGAGACAAGCTGGAACCTGAAACTTGACCAGACTCATTTAG 1062		

Db	100655	CAGCTGTGGTAAAGGTGACTGACTCTGGAGACAGGCTGAAACTGGACCAGACTCATTTAG	100714
QY	1063	AGACAGTAATTCACGACACCAGGAAAGAAATCTTAGTTTAAATGGAGGCTACAGAGGAA	1122
Db	100715	AGACAGTCAATTCGCGCCCGGCAAGGGTCTTAGTTTAAATGGAGGCTACAGAGGAA	100774
QY	1123	ATGAAGGTACCCTTGAAGATCCATCAATGAGAAGACTTTTTTCAGCTACTATCGTCATTGAA	1182
Db	100775	ATGAAGGCACCCCTCAATACATCAATGAGAAGGCTTTTTTCGGCTAAGATAGTCATTGAA	100834
QY	1183	CTGCGCCCTTTAAAGAGCGCAGAGTTGAAGAAATCAATATGAAGACATTTCTAAACTTG	1242
Db	100835	TTGAGACCTTTGAAAGAACACAGAGTAGAAGCTATTCAATATAAAGACATATCTAATCTTG	100894
QY	1243	CCTGAGTTTGAAGATTTGTTAAACAATCAATCAATCAATCAATCAATCAATCAATCAAT	1302
Db	100895	CTTGAGTTTGAAGATTTGTTTGAACAACATTTGAATTTGTAAGCATCAATTTGATGCTAG	100954
QY	1303	CAAAGGCATTTAGACACTCTACTGTGTAGGGTATATCTTTTGTATATAAACAACAGGT	1362
Db	100955	TCAGGCACTTTGTAACCTCTACTGTGTAGGGATTTGTTTGTATATAAATAATCATCT	101014
QY	1363	TTTTGAAATATTTACTGTATGTTGTTGAGCTAAACTTTGAGAAAGAATTTAATTA	1417
Db	101015	ATTT--AAGTAATACTAAATAATTCGTTAAATAAATCTTATGTTATTTTTTTAA	101067
RESULT 11			
LOCUS	AC140205	140215 bp	DNA linear ROD 29-MAY-2004
DEFINITION	Mus musculus BAC clone RP24-344F2 from chromosome 8, complete sequence.		
ACCESSION	AC140205		
VERSION	AC140205.2	GI:47131361	
KEYWORDS	HTG.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.		
AUTHORS	Tomlinson, C., Bielicki, L., Cordum, H. and Haekenson, W.		
TITLE	1 (bases 1 to 140215)		
JOURNAL	The sequence of Mus musculus BAC clone RP24-344F2 Unpublished (2001)		
REFERENCE	2 (bases 1 to 140215)		
AUTHORS	McPherson, J.D. and Waterston, R.H.		
TITLE	Direct Submission		
JOURNAL	Submitted (23-FEB-2003) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
REFERENCE	3 (bases 1 to 140215)		
AUTHORS	Wilson, R.K.		
TITLE	Direct Submission		
JOURNAL	Submitted (12-MAY-2004) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
REFERENCE	4 (bases 1 to 140215)		
AUTHORS	Wilson, R.K.		
TITLE	Direct Submission		
JOURNAL	Submitted (29-MAY-2004) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
COMMENT	On May 12, 2004 this sequence version replaced gi:28475441.		
	----- Genome Center		
	Center: Washington University Genome Sequencing Center		
	Center code: WUGSC		
	Web site: http://genome.wustl.edu		
	Contact: submissions@wustl.edu		
	----- Summary Statistics		
	Center project name: M_BB0344F02		

	NOTICE:		

		/rpt_family="BRV1"	
repeat_region	46796..47083		
repeat_region	/rpt_family="L1"		
repeat_region	47434..47573		
repeat_region	/rpt_family="B2"		
repeat_region	47713..48088		
repeat_region	/rpt_family="L1"		
repeat_region	48495..48643		
repeat_region	/rpt_family="Alu"		
repeat_region	49036..49254		
repeat_region	/rpt_family="MER1_type"		
Query Match	49.7%; Score 759; DB 9; Length 140215;		
Best Local Similarity	77.1%; Pred. No. 1.2e-149;		
Matches 1060; Conservative	0; Mismatches 260; Indels 55; Gaps 9;		
Qy	46	TGATCGCTGCCGTGGTCCGCATTGGGGAAGTCGGATTTTCTTACTCCCAAGGCTATCGCCA	105
Db	81734	TGGTCGGGCCAAGGTAGCCTTGTCALGGCGG-TTTTTTAGCCTTAAGGCCATCGCCA	81792
Qy	106	ACAGA--TCAAGTCCAAAGGGGCTGCAGAAGCTACGCTGGTATTGCCAGATGTGCCAAG	164
Db	81793	ATAAGATTAAGTCCAAAGGCTCCAGAAGCTTCAACCGTACTGCCAGATGTGCCAAAAG	81852
Qy	165	CAGTCCCGGACGAGAAATGGCTTTTAAGTGTCTCATTTGTCTCGAATCTCATCACAGACA	224
Db	81853	CAATGCCGCCACGAGAAATGCTTTTAAGTGTCTCATGTATGTCTAAATCTCATCAAAGACC	81912
Qy	225	CTATTGCTGGCTTCAGAAATCCTCAGCAGTTTATGGATTATTTTCAGAGGAATTCGA	284
Db	81913	CAGTTGCTGGCTTCAGAAAGCCTCAGCAGTTTATACATTATTTTTCAGA-----	81962
Qy	285	AATGACTTTTCTAGAACCTTCTCAGGAGAGCGCTTTGGCACCTAAAAAGGCTCCACAAACA	344
Db	81963	-----GCATTAAGGGCCCCACAACAATT	81988
Qy	345	GTCCTACAAAGAAATACATCAGCCACCGAGAGACATCCACATGAATGCCACTCAGTGGAA	404
Db	81989	GTCACAAATGAATATATCAGCCACCAAGAGCACATCCACATGAATGCTACCGAGTGGAA	82048
Qy	405	ACTCTGACTGNNTTTACTAAGTGGCTGGGCAGAGAGCGCTTGTCAAAAGTGGACGAGACA	464
Db	82049	ANGCTCACCACGPTTTTCCAAAGTGGCTGGGCAGA-AGGGCTTGTGTAAAGTGGGTGAGACA	82107
Qy	465	CCAAAAGGCTGTATATTTCAGTACATAGACAGGSCCCAGAACTATCCGCGGCAACTG	524
Db	82108	CCAAAAGGCTGATACA-TCTGTACATAGACAGAGCCCAGAACTATCTGTGGGCAACTG	82166
Qy	525	GAACT-GGAGAAAGAAAGACGAGACCTTGATGATGAAGAAAAAACTGCCAAATTTAT	583
Db	82167	AAATTAGAAAAAAGAGATGGAAGATCTTCACGATGAAGAAAAAACTGCCNAATTCAT	82226
Qy	584	TGAAGAGCACTGAGAAGAGGCTCTGAAGGGAGGNAAGAGGCTCCCT-ACTTTTACGG	642
Db	82227	TGAGSAGCAGGTGAGAAGAGCGCTTGAAGGGAAAGCAGGAGACACCTGTTTTTTACAG	82286
Qy	643	AA'TTAAAGCAGAGAAATGATAAGAGAAAAAGTCAAGCTTAA'TTGTAGTAAGAGGACATGTA	702
Db	82287	AA'ITAGGCAGAGAAATNAGGAAGAAATTTATGCTCAATCTGTAATAAGGAGTGGGTG	82346
Qy	703	GCTCATCCGGACAAACATCTTCCAAGTCAAGTACTCTGGGACCGAGTGCACTGAAGACGA	762
Db	82347	GCTCAGCAGGAGCTACAAAGTCCAGGTCAAGTCTTTTGGGACCAAGTGCACTGAAGCTGC	82406
Qy	763	TAGGAAGTTCAGCATCAGTGAACGAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAG	822
Db	82407	TGAGGAGCAGACATCCGGGAAATGGAAGGGGTCTTTCACAGAGCTCCTCCCAGCCT---	82463
Qy	823	AAAAGAAGAAAAAGAAATCTGCACTGGATGAAATCATGGAGATTGGAAGAGGAAAAAGAAA	882
Db	82464	CAAGAAGAAGAGATGCTCTGCCCTGGATGAATTTATAGATTAAAGAAGAAAGAGAAA	82523
Qy	883	GAACTGCCCGAACAGACTACTGGCTACGCTGAAATTTATGTGAAAAATTTATAACCAAGA	942

AUTHORS Klein, S. and Gerhard, D.S.
TITLE Direct Submission
JOURNAL Submitted (23-JUN-2004) National Institutes of Health, Xenopus Gene Collection (XGC), National Institute of Child Health and Human Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD 20892-7510, USA
REMARK NIH-MGC Project
COMMENT Contact: XGC help desk
 Email: cgsbbs-remail.nih.gov
 Tissue Procurement: Robert M. Grainger
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: DOE Joint Genome Institute, Walnut Creek, California
 Web site: www.jgi.doe.gov

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAX Plate: 130 Row: b Column: 23
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, similarity but not identity to protein.

FEATURES

Location/Qualifiers
 1..1483
 /organism="Xenopus tropicalis"
 /mol_type="mRNA"
 /db_xref="taxon:8364"
 /clone="MGC:69184 IMAGE:5307170"
 /tissue type="Embryo, Silurana tropicalis, gastrula (stages_10-13)"
 /clone_lib="NICHG_XGC_Emb5"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6.1"
 1..1483
 /gene="MGC69184"
 /db_xref="GeneID:448079"
 59..1228
 /gene="MGC69184"
 /codon_start=1
 /product="MGC69184 protein"
 /protein_id="AAH74606.1"
 /db_xref="GI:49257754"
 /db_xref="GeneID:448079"
 /translation="MGRADFLTPKAIANRIKSLGKLKRWYQMCQKQCRDENGFKCH
 CMSHSHQRLLLSENPOQFDYFSDERTEFLRLRRFQTKRVHNNIYVNEYISHR
 EHVHNNATQWETLDFTKWLRAGYCKVDTPFKWYIQYIDRDPETTRKQEQEKKK
 QDLDDERTAKFIEQVKKGLEVTQNTPTVTELSRONEEEKVAFNLNKGASTSGTAT
 TKSSLSVNLKSAVMGSVKRKDTSOVPKEKKKSLALDEIMSEEEKKKKTERTDWF
 LOPEIVKIVTKLGEKKYKKKAVVKEVIDRYTAVVKLVDSGDKLKLDOHLETVIPA
 PKGRFVLNGYKGFHGLDCINKEFSFATIEIDSGPLKGRKRVGIPYEDISKVA"

ORIGIN

Query Match 46.1%; Score 705; DB 5; Length 1483;
 Best Local Similarity 75.4%; Pred. No. 5,3e-138;
 Matches 908; Conservative 0; Mismatches 285; Indels 12; Gaps 2;
 QY 54 GCCGTGTCGCATGGGGAAGTCGGATTTCTTACTCCCAAGGCTATCCCAACAGGATC 113
 DB 47 GCAGGCTTGGCCATGGGGAAGCAGACTTCCTTACTCCGAAGGCAATCGCAATCGTATC 106
 QY 114 AAGTCCAGGGGCTGCAGAACTAGCTGGTATTGCCAGATGTGCCAGAACAGTGC CGG 173
 DB 107 AAATCGAAAGGACTCCAGAACTCGCATGCTGCTCAAAATGTGTGAGAAACAGTGTCCG 166
 QY 174 GACGAGATGGCTTTAAGTGTCAATGTATGTCCGAATCTCATCAGAGACAATTTGCTG 233
 DB 167 GATGAGATGCTTTAATGCCACTGCATGCTGAGTCGCACCAAGACAATGCTCTCTT 226
 QY 234 GCTTCAGAAAATCCTCAGCAGTTTATGATTTATTTTCAGAGGAATTCGAAATGACTTT 293
 DB 227 GCTTCTGAAAATCCTCAGCAGTTTATGATTTATTTTCAGATGAATTTTCGACAGAAAT 286
 QY 294 CTAGAACTTCTCAGAGAGCGCTTTGGCACTAAAGGCTCCACACAAACATTTGTCTACAC 353

DB 287 TTAGAAATGCTAGGCGTCGGTTTGGAAACCAAGAGATACATATATATTTGTCTACAAT 346
 QY 354 GAATACATCAGCCACCCAGAGACACATCCACATGAATGCCACTCAGTGGGAAACTCTGACT 413
 DB 347 GAATATATCAGCCATAGGGAGCATGTACACATGAATGCCACCCAGTGGGAAACACTTACA 406
 QY 414 GATTTTACTAAGTGGCTGGGCGAGAGAAGGCTTTGTGCAAGTGGACGAGACACCAAAAGGC 473
 DB 407 GACTTCCAAATGGCTTGGCCGAGCAGGTTACTGTAAAGTGGATGAGACTCCAAAGGGC 466
 QY 474 TGGTATATTCAGTACATAGACAGCGGCCACAGAACTATCCGCCGCAACTTGGAACTGGAG 533
 DB 467 TGGTACATTCAGTATATAGACAGATCCAGAAACAATTCGAAACAGCAGGAACAGAA 526
 QY 534 AAAAGAAAAAGCAGGACCTTTGATGATGAAGAAAAAACTGCCAAATTTATTGAGAGACAA 593
 DB 527 AAGAAGAAAAACAAGATCTAGATGATGAAGAGAGAACTGCTAAATTTATTGAGAAACAA 586
 QY 594 GTGAGAAAGCGCTTGGAAAGGAAAGAACAGAGGTCCCTACTTTTACGGAAATTAAGCAGA 653
 DB 587 GTTAAAAAGGGCTTGGAAAGTCACTGAACAGAACACCCCTACCTACACTGAATTTAAGCAGA 646
 QY 654 GAAATGATGAAGAGAAAGTCAAGTTTAAATTTGAGTAAAGAGCATGTAGCTCATCCGGA 713
 DB 647 CAGAAAGGAAAGAAAGTTGCAATTCATTTAAACAAAGAGCA---AGCACTTCAGGA 703
 QY 714 GCAACATCTTCCAAGTCAAGTACTCTGGGACCGAGTGCACCTGAAGACGATAGGAAGTTCA 773
 DB 704 ACAGCTACAACTAAGTCAAGTCTCTTTGTTATCCAATGCACTTAAAGTCTGCTGCTTTATG 763
 QY 774 GCATCAGTGAACGAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAAAGAGAA 833
 DB 764 GGCTCAGTAAAAAGAAAGACACCTCTCAGGTGCAACC-----AAAGGAAAAAGAG 814
 QY 834 AAGAAATCTGCACCTGGATGAAATCATCGAGATTTGAAGAGGAAAAAGAAAGAACTGCCCGA 893
 DB 815 AAGAAATCTGCCCTTGTGATGAGATTATGGAGATGGAGAAAAAGAAAGAAAGACAGAGCGC 874
 QY 894 ACAGACTACTGGCTACAGCCCTGAAATTTATTGTGAAAAATTATAACCAAGAACTCGGAGAG 953
 DB 875 ACAGATTTTGGTTTACAGCCTGAAATTTGTTGTAATAATTTGTTACGAAAAAATTTGGGAGAA 934
 QY 954 AAATATCATAGAAAAAGGCTATTGTTAAGAAAGTAAATTGACAAATATACAGCTGTTGTG 1013
 DB 935 AAATACTACAAGAAAAAGGCTGTAGTAAAGGAAAGTGAATGACAGATACACTGCAAGTTGTG 994
 QY 1014 AAGATGATTGATTCTCGAGACAAGCTGAAACTTTGACCAGACTCATTTTAGAGACAGTAAT 1073
 DB 995 AAATCTGTTGATCTCGAGATAAATTAAGCTGGATCAGACCCCATTTAGAGACAGTTATC 1054
 QY 1074 CCAGCACCCAGAAAAAGAAATTTCTAGTTTTAAATGGAGGCTACAGAGAAATTAAGAGTACC 1133
 DB 1055 CCAGCACCCAGAAAAAGAGAGTTTTTTGTTTTTAAATGGTGTGTTTACAAAGGACATTTTGGCATC 1114
 QY 1134 CTAGAAATCCATCAATGAGAAAGACTTTTTCAGCTACTATCTGCTATTGAAACTGGCCCTTTA 1193
 DB 1115 CTAGATTGTCATCAATGAGAAAGAGCTTTTCGGCTACAAATAGAAAATTTGACTCAGAGCACTTA 1174
 QY 1194 AAAGGCGCAGAGTTGAAGGAATTTCAATATGAAGACATTTTCTAAACTTGTCTGATTTGA 1253
 DB 1175 AAAGGACGGAAGTGGGAAGGCATCCCATATGAAGACATATCCAAAGTTGCTTAGATGTAT 1234
 QY 1254 AAAT 1258
 DB 1235 TAAAT 1239

RESULT 13
 AC141995/c
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-293H7, *** SEQUENCING IN PROGRESS
 AC141995 144118 bp DNA linear HTG 21-MAR-2003
 ***, 48 unordered pieces.

Wed Nov 30 09:40:38 2005

AC141995
 AC141995.1 GI:29135466
 HTG: HTGS_PHASE1
 Rattus norvegicus (Norway rat)
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurionathi; Muroidae; Muridae; Murinae; Rattus.
 1 (bases 1 to 144118)
 Muzny, D. Marie., Metzker, M., Abramson, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D.,
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Bundy, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Loreschewski, L., Loulseghe, H., Lozano, R.J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawhney, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A.,
 Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K.,
 Morris, S., Munidasa, M., Murphy, M., Nair, J., Nankervis, C., Neal, D.,
 Newton, N., Nguyen, N., Norris, S., Nwaokeme, O., Okwuonu, G.,
 Olarunsgaon, A., Pal, S., Parks, K., Pasternak, S., Paul, H.,
 Perez, A., Perez, L., Pfankuch, C., Plopper, F., Poindexter, A.,
 Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E.,
 Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y.,
 Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A.,
 Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S.,
 Scott, G., Shateman, S., Shen, H., Shetty, J., Shvartsbeyn, A.,
 Sisson, I., Sitter, C.D., Smales, D., Sneed, A., Sodergren, E.,
 Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A.,
 Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S.,
 Tingey, A., Trejos, Z., Umami, K., Valas, R., Vera, V., Villasana, D.,
 Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J.,
 Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R.,
 Wooden, H., Wolley, K., Wright, D., Wright, R., Wu, J., Yakub, S.,
 Yen, J., Yoon, H., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X.,
 Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R.,
 Holt, R.A., Smith, H.O., Weinstein, G. and Gibbs, R.A.
 Direct Submission
 Unpublished
 2 (bases 1 to 144118)
 Worley, K.C.
 Direct Submission
 Submitted (21-MAR-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GRGN
 Center clone name: CH230-293H7

Summary Statistics

Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 108995 bases at least Q40
 Consensus quality: 114645 bases at least Q30
 Consensus quality: 119503 bases at least Q20
 Estimated insert size: 114401; sum-of-contigs estimation
 Quality coverage: 2x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently

* consists of 48 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 1394: contig of 1394 bp in length
 * 1395 1494: gap of unknown length
 * 1495 2850: contig of 1356 bp in length
 * 2851 2950: gap of unknown length
 * 2951 3972: contig of 1022 bp in length
 * 3973 4072: gap of unknown length
 * 4073 5309: contig of 1237 bp in length
 * 5310 5409: gap of unknown length
 * 5410 5564: contig of 1155 bp in length
 * 5565 6664: gap of unknown length
 * 6665 8449: contig of 1785 bp in length
 * 8450 8549: gap of unknown length
 * 8550 9903: contig of 1354 bp in length
 * 9904 10003: contig of 1743 bp in length
 * 10004 11746: gap of unknown length
 * 11747 11846: gap of unknown length
 * 11847 13518: contig of 1672 bp in length
 * 13519 13618: gap of unknown length
 * 13619 14970: contig of 1352 bp in length
 * 14971 15070: gap of unknown length
 * 15071 16373: contig of 1303 bp in length
 * 16374 16473: gap of unknown length
 * 16474 18359: contig of 1886 bp in length
 * 18360 18459: gap of unknown length
 * 18460 19821: contig of 1362 bp in length
 * 19822 19921: gap of unknown length
 * 19922 21449: contig of 1528 bp in length
 * 21450 21549: gap of unknown length
 * 21550 23269: contig of 1720 bp in length
 * 23270 23369: gap of unknown length
 * 23370 25710: contig of 2341 bp in length
 * 25711 25810: gap of unknown length
 * 25811 27760: contig of 1950 bp in length
 * 27761 27860: gap of unknown length
 * 27861 29815: contig of 1955 bp in length
 * 29816 31213: contig of 1298 bp in length
 * 31214 31313: gap of unknown length
 * 31314 33042: contig of 1729 bp in length
 * 33043 33142: gap of unknown length
 * 33143 34194: contig of 1052 bp in length
 * 34195 34294: gap of unknown length
 * 34295 36020: contig of 1726 bp in length
 * 36021 36120: gap of unknown length
 * 36121 37671: contig of 1551 bp in length
 * 37672 37771: gap of unknown length
 * 37772 39759: contig of 1988 bp in length
 * 39760 42442: contig of 2583 bp in length
 * 42443 42542: gap of unknown length
 * 42543 44482: contig of 1940 bp in length
 * 44483 44582: gap of unknown length
 * 44583 46351: contig of 1769 bp in length

* 46352	46451: gap of unknown length	
* 46452	48712: contig of 2281 bp in length	
* 48812	gap of unknown length	
* 48813	51252: contig of 2440 bp in length	
* 51352	gap of unknown length	
* 51353	54232: contig of 2880 bp in length	
* 54233	gap of unknown length	
* 54332	gap of unknown length	
* 57136	contig of 2804 bp in length	
* 57137	57236: gap of unknown length	
* 57237	59833: contig of 2597 bp in length	
* 59933	gap of unknown length	
* 59934	62619: contig of 2686 bp in length	
* 62620	62719: gap of unknown length	
* 62720	65087: contig of 2368 bp in length	
* 65088	65187: gap of unknown length	
* 68329	contig of 3142 bp in length	
* 68430	72825: gap of unknown length	
* 72825	contig of 4396 bp in length	
* 72826	72925: gap of unknown length	
* 76196	contig of 3271 bp in length	
* 76197	76296: gap of unknown length	
* 76297	80327: contig of 4031 bp in length	
* 80328	80427: gap of unknown length	
* 83596	contig of 3169 bp in length	
* 83597	83696: gap of unknown length	
* 88053	contig of 4357 bp in length	
* 88153	gap of unknown length	
* 88154	91623: contig of 3470 bp in length	
* 91624	91723: gap of unknown length	
* 91724	96501: contig of 4778 bp in length	
* 96502	96601: gap of unknown length	
* 96602	101193: contig of 4592 bp in length	
* 101194	101293: gap of unknown length	
* 101294	106768: contig of 5475 bp in length	
* 106769	106868: gap of unknown length	
* 106869	111636: contig of 4768 bp in length	
* 111637	111736: gap of unknown length	
* 111737	117842: contig of 6106 bp in length	
* 117843	117942: gap of unknown length	
* 117943	130224: contig of 12282 bp in length	
* 130225	130324: gap of unknown length	
* 130325	144118: contig of 13794 bp in length.	
FEATURES		Location/Qualifiers
source		1..144118
		/organism="Rattus norvegicus"
		/mol_type="genomic DNA"
		/db_xref="taxon:10116"
		/clone="CH230-293H7"
gap		1395..1494
gap		/estimated_length=unknown
		2851..2950
Query Match		44.8%; Score 685; DB 14; Length 144118;
Best Local		Similarity 80.7%; Pred No. 4,6e-134;
Matches		827; Conservative 0; Mismatches 190; Indels 8; Gaps 2;
Qy 439	AAGCCTTGCAAGTGGACGAGACACCAAAAGGCTGGTATATTCAGTACATAGACAGG	498
Db 121995	AAGCCTTGTAAGTGGATGAGACACCAAAAGGCTGGTACATTCAGTACATAGACAG	121936
Qy 499	ACCAGAACTATCCGCGGCACTGGAACTGGAGAAAAGAAAGCAGCCTTGATG	558
Db 121935	ACCAGAACTATACGTCCGCACTGGAACTAGAAAAAAGAAAGCAAGATCTTGACG	121876
Qy 559	ATGAAGAAAAAAGTCCCAATTTATTGAAGACAGTGCAGAAAGAGCGCTGGAAGGGAAGG	618
Db 121875	ATGAAGAAAAAAGTCCCAATTCATTGAGGACAGTGCAGAAAGAGCGCTAGAAAGGAAG	121816
Qy 619	AACAGGAGGTCCTACTTTTACGGAAATTAAGCAGAGAAAAATGATGAAGAGAAAGTCAGT	678
Db 121815	AGCAGGAGCGCCTGCTCTTACAGAGTTTAAGTCGAGAAAAATGATGAAGAAAAAGGTCAGT	121756
Qy 679	TTAATTTGAGTAAGGACGATAGCTCATCCGGAGCAACATCTTCCAAGTCAAGTACTC	738

Db 121755	TCATCTGAATAAAGGAGCGGTAGCTCAGCGGAGCAACAGCTCCAAGTCAAGCTCGC	121696
Qy 739	TGGACCGAGTGCACCTGAAGACCATAGGAAGTTTCAGCATCAGTGAACGAAAAAGAACTTT	798
Db 121695	TGGACCGAGCGCACCTGAAGATGCTGGGAGCACAGCATCCGTGAGCGGAAAGAGTCTT	121636
Qy 799	CCGAGAGCTCAACTCAGTCTAAAGAAAGAAAAAGAAATCTGCACTCGATGAATCA	858
Db 121635	CACAGAGCTCAGCTCAGCCT---GCGAAGAAAGAAAGAAAGTCTGCCCTCGATGAGATCA	121579
Qy 859	TGGAGATTGAAGAGGAAAAAGAAAGAACTCCCGAACAGACTACTGCTACAGCCTCAAA	918
Db 121578	TGGAGATTGAGGAGGAAAAAGAAAGAAAGGACCGACGACAGATTCCTGTTTACAGCTGGGA	121519
Qy 919	TTATTGTGAAAAATTATAACCAAGAAACTGGGAGAGAAATATCATAGAAAAAGGCTATTG	978
Db 121518	TCCTTTGGGAAAAATTATAGGAAGAAAGCTTCGGGAGAAATACCAAGAAAGAGGCTCG	121459
Qy 979	TTAAGGAAGTAATTGACAAATATACAGCTGTTGTGAAGATGATTCCTGGAGACAAGC	1038
Db 121458	TTAAGGAAGTGAATTGACAGATACACAGCTGTGGTAAAGATGACTGACTCCGGAGACAGC	121399
Qy 1039	TGAAACTTGAACAGACTCATTTAGAGACAGTAAATCCAGCACAGGAAAAAGAAATTCAG	1098
Db 121398	TGAACTGGACACAGACTCATTTAGAGACAGTCAATCCAGCCCAAGGAAAGAGTTCAG	121339
Qy 1099	TTTTAAATGGAGGCTACAGAGGAAATGAAGGTACCCCTAGAAATCCATCAATGAGAAAGCTT	1158
Db 121338	TTTTAAACGGAGGCTACAGAGGAAATGAAGGCAACCTCGAGTCCATCAATGAAAAAGACTT	121279
Qy 1159	TTTCAGCTACTGCTCATTTGAAACTGGCCCTTTAAAGAGCAGAGTGTGAAGGAATTC	1218
Db 121278	TTTCAGCCACGATAGTCAATTGAAACCGGACCTTTGAAAGGGCGCAGACTTGAAGGTATTC	121219
Qy 1219	AATATGAAGACATTTCTAAACTTTCCTGAGTGTGAAATTTGTTAAACAATACATTAAAAAT	1278
Db 121218	AGTATGAAGACATATCTAAACTTTCCTGAGTGTGAAATTTGTTAAACAATACATTGAAAT	121159
Qy 1279	CTTAAAGCATCAAAATGGTGTTCGCCAAGCATATTAGAGACTCTACTGTGTAGGGTATA	1338
Db 121158	TGTAAGCATCAAAATGATGTAGCCACGCACTGTGTAACCTCTACTGTGTAGGGGATT	121099
Qy 1339	TTCTTTGTATATAAACAACACAGGTTTTTGAATAATATTACTGTAGTGTTCAGCTAAAC	1398
Db 121098	TGTTTAT-----AAAGAAAGTCATCTATTTAAATAACCACTGGATAGCTGTTGGGTAAAT	121044
Qy 1399	TTTGAGAGAAATTTAATTATGTCTCATGAGGTATCAAACTATGTAAATTTTGTCTCTGTTA	1458
Db 121043	TTATATAAATCTAATTTTTTTTAAAGTGCATTTTATCTTTGTTCTGTTTCTGATGG	120984
Qy 1459	TTTTT 1463	
Db 120983	ATTTT 120979	
RESULT 14	AC114615/c	
LOCUS	AC114615	180882 bp DNA linear HTG 09-MAR-2004
DEFINITION	Mus musculus clone RP24-82119, WORKING DRAFT SEQUENCE, 4 ordered	
	pieces.	
ACCESSION	AC114615	
VERSION	AC114615.5	GI:44886278
KEYWORDS	HTG; HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;	
	Sciurognathi; Muridae; Muridae; Murinae; Mus.	
REFERENCE	1 (bases 1 to 180882)	
AUTHORS	Birren,B., Nusbaum,C. and Lander,E.	
TITLE	Mus musculus, clone RP24-82119	
JOURNAL	Unpublished	

REFERENCE
AUTHORS

2 (bases 1 to 180882)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Bloom,I., Boguslavskiy,I.,
 Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
 Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A.,
 Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., Lacroque,K., Lamazares,R.,
 Landers,T., Lehoczyk,J., Levine,R., Lindblad-Toh,K., Liu,G.,
 MacLean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,
 McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
 Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE
JOURNAL

Submitted (10-MAR-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE
AUTHORS

3 (bases 1 to 180882)
 Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
 Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
 Boguslavskiy,I., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
 Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
 Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
 Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
 Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
 Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
 Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
 MacDonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
 Meldrum,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
 Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
 O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
 Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
 Roman,J., Schauer,S., Schupack,R., Seaman,S., Severy,P., Smith,C.,
 Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
 Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M.,
 Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
 Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE
JOURNAL

Submitted (09-MAR-2004) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 2, 2004 this sequence version replaced gi:24182248.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

COMMENT

 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence.submissions@genome.wi.mit.edu

 Project Information

 Center project name: L24315
 Center clone name: 82_1_19

 Summary Statistics

 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 191018 bases at least Q40
 Consensus quality: 191303 bases at least Q30
 Consensus quality: 191388 bases at least Q20
 Insert size: 178000; agarose-
 Insert size: 191574; sum-of-contigs
 Quality coverage: 12.2 in Q20 bases; agarose-
 Quality coverage: 11.3 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 89963: contig of 89963 bp in length
 * 89964 90063: gap of 100 bp
 * 90064 94560: contig of 4497 bp in length
 * 94561 94660: gap of 100 bp
 * 94661 169601: contig of 74941 bp in length
 * 169602 169701: gap of 100 bp
 * 169702 180882: contig of 11181 bp in length.
 Location/Qualifiers
 1..180882
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /clone="RP24-82119"
 /clone_lib="RPC1-24 Male Mouse BAC"
 1..89963
 /note="assembly_fragment
 clone_end:SP6
 vector_side:left"
 89964..90063
 /estimated_length=100
 90064..94560
 /note="assembly_fragment"
 94561..94660
 /estimated_length=100
 94661..169601
 /note="assembly_fragment"
 169602..169701
 /estimated_length=100
 169702..180882
 /note="assembly_fragment
 clone_end:17
 vector_side:right"

 Query Match 44.1%; Score 673.2; DB 14; Length 180882;
 Best Local Similarity 78.2%; Pred. No. 1.4e-131;
 Matches 823; Conservative 0; Mismatches 223; Indels 6; Gaps 1;

 QY 439 AAGGCTTGTCAAAGTGGACGAGACACCAAAAGGCTGTATATTCAGTACATAGACAGG 498
 |||||
 Db 46867 AAGGCTTGTCAAAGTGGATGAGACACCGAAGGCTGGTACATTCAGTACATAGACAGAG 46808
 |||||
 QY 499 ACCGAGAACTATCCCGCGCACTGGAACTCGAGAAAAAGAAAAAGCAGGACCTTGATG 558
 |||||
 Db 46807 ACCGAGAACTATCCCGCGCACTGGAACTAGAAAAAGAAAGCAAGATCTGGACG 46748
 |||||
 QY 559 ATGAGAAAAAACTCCCAATTTATTTAGAGACAGTGAAGAGCGCTGGGAAGGGAAGG 618
 |||||
 Db 46747 ATGAGAAAAAACTGCCAAGTTTCATTTAGAGACAGTGAAGAGCGCTGGGAAGGGAAG 46688
 |||||
 QY 619 AACAGGAGGTCCTCTACTTTTACGGAAATTAAGCAGAGAAAAATGATGAAGAGAAAGTCACGT 678
 |||||
 Db 46687 AGCAGAGACACCTGTTTTTACAGAACTTAGCCGAGAAATGAGGAAGAAAAAGTTACGT 46628
 |||||
 QY 679 TTAATTTAGTAAAGAGCATGTAGCTCATCCGAGCAACATCTTCCAAAGTCAGTACTC 738
 |||||
 Db 46627 TCAATCTGAATAAAGGAGCGGTGGCTCAGCGGAGCTCAACACATCCAAAGTCAAGCTCTT 46568
 |||||
 QY 739 TGGGACCGAGTCGACTGAGAGCATAGGAAGTTTCAGCATCAGTGAACCGAAAGAAATCTTT 798
 |||||
 Db 46567 TGGGACCAAGTGCACCTGAAAGCTGCTGGGAGCGGCAGCATCCCGGAAACCGAAAGAGCTTT 46508
 |||||
 QY 799 CCCAGAGCTCAACTCAGTCTTAAGAAAAAGAAAAAGAAATCTGCATCTGGATGAATCA 858
 |||||

Db	46507	CACAGAGCTCCGCCAGGCT-----GCGAAGAAAGAAAGTGCGCCCTGGATGAGATCA	46454
QY	859	TGGAGATTGAAGAGGAAAGAAAGAACTGCCCGAACAGACTACTCGGTACAGCCTGAAA	918
Db	46453	TGGAGCTCGAAGAGGAAAGAAAGAAAGCGCACGACGCGCTGGTTACAGCGGGGA	46394
QY	919	TTATTGTGAATTATACCAAGAACTCGGGAGAGAAATATCATTAAGAAAAAGGCTATTG	978
Db	46393	TGCTTGTGAAATATATACGAGAAAGCTTGGGGAGAAATATCACAGAAAGAAAGGGTCG	46334
QY	979	TTAAGGAGTATTCACAAATATACAGCTGTGTGCAAGTATGATTCATTCTGCGAGACRAGC	1038
Db	46333	TTAAGGAGTATTCACAGGTTACACAGCTGTGTGTAAGATGACTACTCTGGAGACAGC	46274
QY	1039	TGAACCTTGACAGACTCATTTAGAGACAGTAATTTCCAGCACGAGAAAGAAATTTCTAG	1098
Db	46273	TGAACCTTGACAGACTCATTTAGAGACAGTATTTCCGCCCGGGGAAAGGGTCTTAG	46214
QY	1099	TTTTAAATGGAGGCTACAGAGAAATGAAGGTACCCCTAGAAATCCATCAATGAGAAGACTT	1158
Db	46213	TTTTAAATGGAGGCTACAGAGAAATGAAGGCACCTCTCGAATCCATCAATGAGAAGGCTT	46154
QY	1159	TTTCAGCTACTATCTCTAATGAACTGGCCCTTTTAAAGGAGCGAGCTTGAAGGAATTC	1218
Db	46153	TTTCAGCCAGCATAGTCTATGAACTGGACCTTTTGAAGGAGCGAGAGTTGAAGGTATTC	46094
QY	1219	AATATGAAGACATTTCTAACTTGCTGAGTTGAAATTTGTTAAACATACATTTAAAT	1278
Db	46093	AATATGAAGACATATCTAACTTGCTTGAGTTGAAATTTGATTAACAACATTTGAAAC	46034
QY	1279	CTTAAAGCATCAATTTGGTTCGCAAGGCATTAAGACTCTACTGTTAGGGTATA	1338
Db	46033	TGTGAAGCATCAATTTGGTTCGCAAGGCCTGTGTAACCTCTACTGTTAGGGGATT	45974
QY	1339	TTCTTTTGTATTAACAACACAGGTTTTCGAAATATTAAGTATGTTGTTTCAGCTAAC	1398
Db	45973	TGTTTTGTATTAACAAAAAATAATATCTATTTAAATTAAGTATGTTGTTTCAGCTAAC	45914
QY	1399	TTTGAGAAGATTTAAATATGTTCTCATGAGGTATCAAACTATGTAATTTTGTCCTTGTTA	1458
Db	45913	TTATAATAATCTATGTTTATTTTAAAGTGCATAACTGTTATTTTATCTTGTTCGGTT	45854
QY	1459	TTTTTGTTCCTTCTTAATTTACTTTGATGAGT	1490
Db	45853	TCCCTGATGGATTTTATACCATCATTTAAAGAAAT	45822
RESULT 15			
AC124010			
LOCUS			
DEFINITION	AC124010 182400 bp DNA linear HTG 25-AUG-2002		
ACCESSION	Mus musculus chromosome UNK clone RP23-119N4, WORKING DRAFT		
VERSION	SEQUENCE, 6 unordered pieces.		
KEYWORDS	AC124010.2 GI:22475963		
SOURCE	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.		
ORGANISM	Mus musculus (house mouse)		
REFERENCE	Mus musculus		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
JOURNAL	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;		
AUTHORS	Sciurognathi; Murioidea; Muridae; Murinae; Mus.		
TITLE	1 (bases 1 to 182400)		
REFERENCE	McPherson,J.D. and Waterston,R.H.		
AUTHORS	The sequence of Mus musculus clone		
TITLE	Unpublished		
REFERENCE	2 (bases 1 to 182400)		
AUTHORS	McPherson,J.D. and Waterston,R.H.		
TITLE	Direct Submission		
REFERENCE	Submitted (06-JUN-2002) Genome Sequencing Center, 4444 Forest Park		
AUTHORS	Parkway, St. Louis, MO 63108, USA		
TITLE	3 (bases 1 to 182400)		
REFERENCE	McPherson,J.D. and Waterston,R.H.		
AUTHORS	Direct Submission		
TITLE			

JOURNAL	Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park		
COMMENT	Parkway, St. Louis, MO 63108, USA		
	On Aug 25, 2002 this sequence version replaced gi:21327639.		
	----- Genome Center -----		
	Center: Washington University Genome Sequencing Center		
	Center code: WUGSC		
	Web site: http://genome.wustl.edu/gsc/index.shtml		
	Contact: submissions@wustl.wustl.edu		
	----- Project Information -----		
	Center project name: M.BA0119N04		
	----- Summary Statistics -----		
	Sequencing vector: M13; 0%		
	Chemistry: Dye-primer ET; 0% of reads		
	Chemistry: Dye-terminator Big Dye; 100% of reads		
	Assembly program: Phrap; version 0.990319		
	Consensus quality: 183528 bases at least Q40		
	Consensus quality: 184210 bases at least Q30		
	Consensus quality: 184756 bases at least Q20		
	Insert size: 197000; agarose-fp		
	Quality coverage: 9.53 in Q20 bases; agarose-fp		
	Quality coverage: 9.61 in Q20 bases; sum-of-contigs		

	* NOTE: This is a 'working draft' sequence. It currently		
	* consists of 6 contigs. The true order of the pieces		
	* is not known and their order in this sequence record is		
	* arbitrary. Gaps between the contigs are represented as		
	* runs of N, but the exact sizes of the gaps are unknown.		
	* This record will be updated with the finished sequence		
	* as soon as it is available and the accession number will		
	* be preserved.		

	* 1 11868: contig of 11868 bp in length		
	* 11869: gap of unknown length		
	* 11969: contig of 16938 bp in length		
	* 28907: gap of unknown length		
	* 29007: contig of 43767 bp in length		
	* 72773: gap of unknown length		
	* 72774: 72873: gap of unknown length		
	* 124927: 125026: gap of 52053 bp in length		
	* 125027: 181902: gap of unknown length		
	* 125027: 181902: contig of 56876 bp in length		
	* 181903: 182002: gap of unknown length		
	* 182003: 182400: contig of 398 bp in length.		
FEATURES	Location/Qualifiers		
Source	1..182400		
	/organism="Mus musculus"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:10090"		
	/chromosome="UNK"		
	/clones="RP23-119N4"		
misc_feature	1..11868		
	/note="assembly_name:Contig86"		
gap	11869..11968		
	/estimated_length=unknown		
misc_feature	11969..28906		
	/note="assembly_name:Contig87"		
gap	28907..29006		
	/estimated_length=unknown		
misc_feature	29007..72773		
	/note="assembly_name:Contig88"		
gap	72774..72873		
	/estimated_length=unknown		
misc_feature	72874..124926		
	/note="assembly_name:Contig89"		
gap	124927..125026		
	/estimated_length=unknown		
misc_feature	125027..181902		
	/note="assembly_name:Contig90"		
gap	181903..182002		
	/estimated_length=unknown		
misc_feature	182003..182400		

/note="assembly_name:Contig16"

ORIGIN

Query Match 44.1%; Score 673.2; DB 14; Length 182400;
Best Local Similarity 78.2%; Pred. No. 1.4e-131;
Matches 823; Conservative 0; Mismatches 223; Indels 6; Gaps 1;

QY 439 AAGGCTTGCAAGTGGACGAGACACCAAAAGGCTGGTATATTCAGTACATAGACAGG 498
DB 168546 AAGGCTTGTAAGTGGATGAGACACCGAAAGGCTGGTACATTGATACATAGACAG 168605
QY 499 ACCGAGAACTATCCGCCGCACTGGAACTGGAGAAAAGAAAAGCAGGACCTTGATG 558
DB 168606 ACCGAGAACTATCCGCGCACTGGATTTAGAAAAGAAAAGCAAGATCTGGAGC 168665
QY 559 ATGAGAAAAAATGCAAAATTTATGAAAGAGCAAGTGAAGAGAGCCCTGGAGGGAAGG 618
DB 168666 ATGAGAAAAAATGCAAAATTTATGAGGAGCAGGTGAGAGAGCCCTGGAGGGAAG 168725
QY 619 AACAGAGGTCCCTACTCTTTTACGGATTAAGCAGAGAAAATGATGAAGAGAAAGTCAGT 678
DB 168726 AACAGAGACACTGTTTTTACAGAACTTAGCCGAGAAAATGAGGAGAAAAGTTACGT 168785
QY 679 TTAATTTGAGTAAAGGAGCATGTAGCTCATCCGGAGCAACATCTTCCAAGTCAAGTACTC 738
DB 168786 TCATCTGATTAAGGAGCGGCTGAGCGGAGCTACACATCCAGTCAAGTCTTT 168845
QY 739 TGGGACCGAGTCACCTGAAGAGATAGGAAGTTGAGCATCAGTGAAACGAAAAGAAATCTTT 798
DB 168846 TGGGACCAAGTCACCTGAAGTCTGGGAGCGCAGCATCCGGGAAAACGGAAGAGTCTTT 168905
QY 799 CCAGAGCTCACTCAGTCTAAGAAAAGAAAAGAAAAGAAATTCGCACTGGATGAAATCA 858
DB 168906 CACAGAGCTCCGCCCAAGCT-----GCGAAGAGAAAGAAAGTGGGCTGGATGAGATCA 168959
QY 859 TGGAGATTGAAGAGGAAAAGAAAAGAACTGCCGCAACAGACTACTGGCTACAGCCCTGAAA 918
DB 168960 TGGAGCTCGAAGAGGAAAAGAAAAGAACCGCAGCAGACAGCGCTGTTACAGCCGGGA 169019
QY 919 TTAATTTGAAAATTTAACCAGAAAATCTGGGAGAGAAAATATCATAGAAAAGGCTATTG 978
DB 169020 TCGTTGTGAAAATTTAACCAGAAAGCTTTGGGAGAGAAATATCACAGAAAAGGGGTCG 169079
QY 979 TTAAGGAAGTAAATTGACAAATATACAGCTGTTGTGAAAGATGATTGATCTGGAGACAAGC 1038
DB 169080 TTAAGGAAGTAAATTGACAGTACACAGCTGTGGTAAAGATGATGACTCTGGAGACAGC 169139
QY 1039 TGAACCTTGACAGACTCATTTTAGACACAGTAAATCCAGCACCAGGAAAAGAAATCTTAG 1098
DB 169140 TGAACCTTGGACAGACTCATTTTAGACACAGTCAATCCGGCCCGGGGAAAAGGGTTCTAG 169199
QY 1099 TTTTAAATGGAGGCTACAGAGGAAATGAAGGTACCCCTAGAACTCAATGAGAAGACTTT 1158
DB 169200 TTTTAAATGGAGGCTACAGAGGAAATGAAGGCACTCTCGAATCCATCAATGAGAAGCTTT 169259
QY 1159 TTTTCAGCTACTTCGTCATTGAAATCGGCCCTTTTAAAGGACGACAGTTGAAGGAATTC 1218
DB 169260 TTTTCAGCCACGATAGTCAATTTGAAAATCGGACCTTTTGAAGGACGACAGTTGAAGGTATT 169319
QY 1219 AATATGAAGACATTTCTAAACTTGCCTGAGTTTGAAGAAATTTGTTAAACAATACATTAATA 1278
DB 169320 AATATGAAGACATTTCTAAACTTGCCTGAGTTTGAAGAAATTTGTTAAACAATACATTAATA 169379
QY 1279 CTTTAAAGCATCAATTTGGTGTTCGCAAGGCATTTATGAGACTCTACTGTGTAGGGTATA 1338
DB 169380 TGTGAAGCATCAATTTGGTGTAGCAAGGCACTGTGTAACTCTACTGTGTAGGGGATT 169439
QY 1339 TCTTTTGTATAAAAACAAACAGGTTTTTGAAGATATTTACTGTATAGTTGTTTCAAGCTAAAC 1398
DB 169440 TGTTTTGTATAAAAACAAACAAATCATCTCTATTTTAAATACTAGTGAATAGTTGGGTAAAT 169499
QY 1399 TTTGAGAGAAATTTAATATGCTCTCAGGATCAAACTATGTAATTTTGTCTTGTGA 1458

Db 169500 TTATAATAAAATCTATGTTTTTTTTTAAGTGTCATAACTGTTATTTATCTTTGTTCCGTT 169559
QY 1459 TTTTGTGTTCCCTTTGTAATTTTACTTTGATGACT 1490
Db 169560 TCCCTGATGGATTTTTTACCATCATTAAGAAT 169591

Search completed: November 26, 2005, 19:10:38
Job time : 8131.42 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 12:48:57 ; Search time 980.087 Seconds
(without alignment)
10390.557 Million cell updates/sec

Title: US-09-555-529-1
Perfect score: 1528
Sequence: 1 ctgaatcagcgccgctg.....atgtattataaaaaaaaaa 1528

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq 21.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*
14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1249.8	81.8	1296	2	AAX85549 cDNA of a
2	1002	65.6	1002	2	AAX85552 Probe der
3	990.6	64.8	1390	2	AAX85570 cDNA of a
4	986.4	64.6	1458	2	AAG79936 Murine Ki
5	653.4	42.8	1002	2	AAX85551 cDNA of a
6	531.6	34.8	679	13	Adq56943 Novel can
7	490.8	32.1	591	4	Aah34095 Human col
8	475	31.1	1102	2	AAX85550 cDNA of a
9	380.6	24.9	1372	12	Adp22451 Sea-squir
10	303.8	19.9	398	13	AcF87901 Human SIR
11	298.6	19.5	1241	4	ABL29199 Drosophil
12	284	18.6	300	3	AAa01559 Human col
13	247.2	16.2	3313	4	ABL29198 Drosophil
14	224.4	14.7	1459	3	AAC47248 Arabidops
15	209.4	13.7	4394	4	ABL29184 Drosophil
16	209.4	13.7	4582	4	ABL18194 Drosophil
17	204.8	13.4	1713	13	Adt17629 Plant cDN
18	204.8	13.4	1780	3	AAa52591 cDNA enco
19	204.8	13.4	1780	3	AAa52590 cDNA enco

20	201.6	13.2	1614	13	AD084255
21	201.6	13.2	1713	13	Adx47874 Plant ful
22	201.6	13.2	1764	3	AAa52589 cDNA enco
23	136.4	8.9	1032	3	Aac48794 Arabidops
24	126.4	8.3	549	3	Aac94929 Cat flea
25	100.2	6.6	270	6	ABL73715 Corn tass
26	94.2	6.2	696	3	AAC38477 Arabidops
27	77.4	5.1	431	3	AAC38801 Arabidops
28	60	3.9	60	6	ABn33592 Human spl
29	59.4	3.9	12578	4	AAa46660 Tumour su
30	58	3.8	5935	4	AAa45427 Chemicall
31	56.8	3.7	2985	4	ABl29210 Drosophil
32	55.2	3.6	11422	6	ABK39937 Human che
33	55.2	3.6	11422	6	ABl32219 Human inn
34	55	3.6	2000	8	ADA71938 Rice gene
35	54.4	3.6	65	6	ABn51894 Mouse spl
36	54	3.5	15951	6	ABl33680 Human inn
37	54	3.5	15951	6	ABl34580 Human met
38	54	3.5	15951	6	ABl70373 Chemicall
39	54	3.5	15951	7	ADs99841 Bisulphit
40	53.8	3.5	6831	6	ABl33486 Human inn
41	53.4	3.5	6013	6	ABK31360 Signal tr
C 42	53.4	3.5	6013	6	AAa61264 Human gen
C 43	53	3.5	8056	8	ABZ10246 Haematopo
44	52.6	3.4	2000	11	ACL37108 Rice stre
45	52.6	3.4	113515	6	ABl34175 Human inn

ALIGNMENTS

RESULT 1
AAX85549
ID AAX85549 standard; cDNA; 1296 BP.
XX
AC AAX85549;
XX
DT 07-SEP-1999 (first entry)
XX
DE cDNA of a gene coding for the human kin17 protein.
XX
KW Human; kin17 protein; cell proliferation; fertility;
KW hyperproliferative disease; protein interaction; curved DNA;
KW HIV replication; HIV integration; repair enzyme; as.
XX
OS Homo sapiens.
XX
PN FR2772046-A1.
XX
PD 11-JUN-1999.
XX
PF 09-DEC-1997; 97FR-00015536.
XX
PR 09-DEC-1997; 97FR-00015536.
XX
(COMS) COMMISSARIAT ENERGIE ATOMIQUE.
XX
Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
WPI; 1999-359999/31.
XX
PT New DNA coding for human kin17 protein - useful for controlling cell
proliferation or fertility.
XX
PS Claim 1; Page 30-31; 69pp; French.
XX
CC The present sequence encodes a human kin17 protein. The mammalian kin17
protein is useful for preparing a medicament for controlling cell
proliferation or for controlling fertility. The medicaments can also be
used to treat hyperproliferative diseases. Fragments between amino acids
55 and 235 (preferably between amino acids 129 and 228) of a mammalian
kin17 protein are useful for regulating the interaction between proteins
and curved DNA. The fragment can be used to block replication of HIV or

CC its integration into the human genome or to target repair enzymes to
CC curved DNA sites. Expression vectors for kin17 can be used for
CC controlling cell proliferation

SQ Sequence 1296 BP; 453 A; 240 C; 310 G; 293 T; 0 U; 0 Other;

Query Match 81.8%; Score 1249.8; DB 2; Length 1296;
Best Local Similarity 99.8%; Pred. No. 7.1e-292;
Matches 1262; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 40 AGAAAGTATGCTCGCTGCGTGGCGCATCGGGAAGTCGGATTTCTTACTCCCAAGGCTA 99
DB |||||
QY 32 AGAAAGTATGCTCGCTGCGTGGCGCATCGGGAAGTCGGATTTCTTACTCCCAAGGCTA 91
DB |||||
QY 100 TCGCCAACAGGATCAAGTCCAAAGGGGCTGCAGAGCTACGCTGGTATTGCCAGATGTGCC 159
DB |||||
QY 92 TCGCCAACAGGATCAAGTCCAAAGGGGCTGCAGAGCTACGCTGGTATTGCCAGATGTGCC 151
DB |||||
QY 160 AGAAGCAGTCCGGGACGAGAAATGGCTTTAAGTGTCATGTATGTCCGAATCTCATCAGA 219
DB |||||
QY 152 AGAAGCAGTCCGGGACGAGAAATGGCTTTAAGTGTCATGTATGTCCGAATCTCATCAGA 211
DB |||||
QY 220 GACAACTATTGCTGCTGCCTTCAGAAATCCTCAGCAGTTTATGGATTATTTTCAGAGGAAT 279
DB |||||
QY 212 GACAACTATTGCTGCTGCCTTCAGAAATCCTCAGCAGTTTATGGATTATTTTCAGAGGAAT 271
DB |||||
QY 280 TCCGAAATGACTTTCTAGAACTTCTCAGGAGACGCTTTGGCACTAAAAGGGTCCACAACA 339
DB |||||
QY 272 TCCGAAATGACTTTCTAGAACTTCTCAGGAGACGCTTTGGCACTAAAAGGGTCCACAACA 331
DB |||||
QY 340 ACATTGTCTACAACGAATACATCAGCCACCGAGAGACATCCACATGAATGCCATCACTAGT 399
DB |||||
QY 332 ACATTGTCTACAACGAATACATCAGCCACCGAGAGACATCCACATGAATGCCATCACTAGT 391
DB |||||
QY 400 GGGAACTCTGACTGATTTTACTAGTGTGCTGGGAGAGAGGCTTGTGCAAGTGGACG 459
DB |||||
QY 392 GGGAACTGTGACTGATTTTACTAGTGTGCTGGGAGAGAGGCTTGTGCAAGTGGACG 451
DB |||||
QY 460 AGACACCAAAAGGCTGGTATTTTACAGTACATAGACAGGAGCCAGAAACTATCCGCCGCGC 519
DB |||||
QY 452 AGACACCAAAAGGCTGGTATTTTACAGTACATAGACAGGAGCCAGAAACTATCCGCCGCGC 511
DB |||||
QY 520 AACTGGAATCGAGAAAAAGAAAGAGCAGGACCTTGATGATGAAGAAAAAACTGCCAAT 579
DB |||||
QY 512 AACTGGAATCGAGAAAAAGAAAGAGCAGGACCTTGATGATGAAGAAAAAACTGCCAAT 571
DB |||||
QY 580 TTATTGAAGAGCAAGTGAAGAGAGGCTTGAAGGGAAGCAAGCAGGAGTCCCTACTTTTA 639
DB |||||
QY 572 TTATTGAAGAGCAAGTGAAGAGAGGCTTGAAGGGAAGCAAGCAGGAGTCCCTACTTTTA 631
DB |||||
QY 640 CGGAATTAAGCAGAGAAAAATGATGAAGAGAAAGTCAACGTTTAAATTTGATGAAGAGCAT 699
DB |||||
QY 632 CGGAATTAAGCAGAGAAAAATGATGAAGAGAAAGTCAACGTTTAAATTTGATGAAGAGCAT 691
DB |||||
QY 700 GTAGCTCATCCGGAGCAACATCTTCCAGTCAAGTACTCTGGGACCGAGTGACCTGAAGA 759
DB |||||
QY 692 GTAGCTCATCCGGAGCAACATCTTCCAGTCAAGTACTCTGGGACCGAGTGACCTGAAGA 751
DB |||||
QY 760 CGATAGGAAGTTTCAGCATCAGTGAACGAAAGAAATCTTCCAGAGCTCAACTCAGTCTTA 819
DB |||||
QY 752 CGATAGGAAGTTTCAGCATCAGTGAACGAAAGAAATCTTCCAGAGCTCAACTCAGTCTTA 811
DB |||||
QY 820 AAGAAAAGAGAAAAAGAAATCTGCACCTGGATGAATCATGGAGATTGAAGAGGAAAAAGA 879
DB |||||
QY 812 AAGAAAAGAGAAAAAGAAATCTGCACCTGGATGAATCATGGAGATTGAAGAGGAAAAAGA 871
DB |||||
QY 880 AAAGAACTGCCGAGACAGACTACTGGCTACAGCTGAAATTTATGTGAAATTTAAACCA 939
DB |||||
QY 872 AAAGAACTGCCGAGACAGACTACTGGCTACAGCTGAAATTTATGTGAAATTTAAACCA 931
DB |||||
QY 940 AGAAACTGGGAGAGAAATATCATAGAAAAAGGCTATTGTTAAGGAAGTAAATTGACAAT 999
DB |||||
QY 932 AGAAACTGGGAGAGAAATATCATAGAAAAAGGCTATTGTTAAGGAAGTAAATTGACAAT 991
DB |||||

QY 1000 ATACAGCTGTTGTGAAGATGATTGATTTCTGGAGACAAAGCTGAAACTTTGACAGACTCATT 1059
DB |||||
QY 992 ATACAGCTGTTGTGAAGATGATTGATTTCTGGAGACAAAGCTGAAACTTTGACAGACTCATT 1051
DB |||||
QY 1060 TAGAGACAGTAATTTCCAGCACCCAGGAAAAGAAATCTAGTTTAAATGGAGGCTACAGAG 1119
DB |||||
QY 1052 TAGAGACAGTAATTTCCAGCACCCAGGAAAAGAAATTTCTAGTTTAAATGGAGGCTACAGAG 1111
DB |||||
QY 1120 GAAATGAAGGTACCCCTAGATCCATCAATGAGAAAGACTTTTTCAGCTACTATCGTCATTG 1179
DB |||||
QY 1112 GAAATGAAGGTACCCCTAGATCCATCAATGAGAAAGACTTTTTCAGCTACTATCGTCATTG 1171
DB |||||
QY 1180 AAACTGGCCCTTTAAAAGGACGACAGATGAAGGAATTCATATGAAGACATTTCTAAAC 1239
DB |||||
QY 1172 AAACTGGCCCTTTAAAAGGACGACAGATGAAGGAATTCATATGAAGACATTTCTAAAC 1231
DB |||||
QY 1240 TTGCTGAGTTTGAAGAAATTTGTTAAACAATA-CATTAAGTCTTAAAGCATCAAAATTTGGTG 1298
DB |||||
QY 1232 TTGCTGAGTTTGAAGAAATTTGTTAAACAATACCTTTAAATCTTAAAGCATCAAAATTTGGTG 1291
DB |||||
QY 1299 TTCGC 1303
DB |||||
QY 1292 TTCGC 1296
DB |||||

RESULT 2

AA85552
ID AA85552 standard; cDNA; 1002 BP.

AC AA85552;

DT 07-SEP-1999 (first entry)

DE Probe derived from cDNA of a gene coding for the human kin17 protein.

Human; kin17 protein; cell proliferation; fertility;

hyperproliferative disease; protein interaction; curved DNA;

HIV replication; HIV integration; repair enzyme; probe; ss.

OS Homo sapiens.

PN FR2772046-A1.

PD 11-JUN-1999.

PF 09-DEC-1997; 97FR-00015536.

PR 09-DEC-1997; 97FR-00015536.

PA (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

PI Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;

PS WPI; 1999-359999/31.

PT New DNA coding for human kin17 protein - useful for controlling cell

proliferation or fertility.

CC Claim 7; Page 32-33; 69pp; French.

The present sequence is derived from nucleotides 207-1208 of AA85549,
and is a probe used to isolate human kin17 cDNA. The mammalian kin17
protein is useful for preparing a medicament for controlling cell
proliferation or for controlling fertility. The medicaments can also be
used to treat hyperproliferative diseases. Fragments between amino acids
55 and 235 (preferably between amino acids 129 and 228) of a mammalian
kin17 protein are useful for regulating the interaction between proteins
and curved DNA. The fragment can be used to block replication of HIV or
its integration into the human genome or to target repair enzymes to
curved DNA sites. Expression vectors for kin17 can be used for
controlling cell proliferation

SQ	Sequence 1002 BP; 374 A; 179 C; 235 G; 214 T; 0 U; 0 Other;	
Query Match	65.6%; Score 1002; DB 2; Length 1002;	
Best Local Similarity	100.0%; Pred. No. 5.6e-232;	
Matches 1002; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
QY	215 TCAGAGACAACATATTGCTGGCTTCAGAAAATCCTCAGCAGATTTATGGATTATTTTTCAGA 274	
DB	1 TCAGAGACAACATATTGCTGGCTTCAGAAAATCCTCAGCAGATTTATGGATTATTTTTCAGA 60	
QY	275 GGAATTCGGAATGACTTTCTAGAACTTCTCAGGAGACGCTTTGGCACTAAAGGGTCCA 334	
DB	61 GGAATTCGGAATGACTTTCTAGAACTTCTCAGGAGACGCTTTGGCACTAAAGGGTCCA 120	
QY	335 CAACAACATTGTCTACAAGCAATACATCAGCCACCGAGAGCACTCCACATGAATGCCAC 394	
DB	121 CAACAACATTGTCTACAAGCAATACATCAGCCACCGAGAGCACTCCACATGAATGCCAC 180	
QY	395 TCAGTGGGAACTCTGACTGATTTTACTAAGTGGCTGGGCAGAGAAGGCTTGTGCAAAAGT 454	
DB	181 TCAGTGGGAACTCTGACTGATTTTACTAAGTGGCTGGGCAGAGAAGGCTTGTGCAAAAGT 240	
QY	455 GGACGAGACACCAAAAGGCTGTATATTTCAGTACATAGACAGGACCCAGAAATCTATCCG 514	
DB	241 GGACGAGACACCAAAAGGCTGTATATTTCAGTACATAGACAGGACCCAGAAATCTATCCG 300	
QY	515 CCGCAACTGGAACTGGAGAAAAAGAAAGCAGACCTTGATGATGAAGAAAAAATCTGC 574	
DB	301 CCGCAACTGGAACTGGAGAAAAAGAAAGCAGACCTTGATGATGAAGAAAAAATCTGC 360	
QY	575 CAAATTTATTGAAGAGCAAGTCAGAGAGCGCTGGAAGGGAAGGAGGAGGTCCTTAC 634	
DB	361 CAAATTTATTGAAGAGCAAGTCAGAGAGCGCTGGAAGGGAAGGAGGAGGTCCTTAC 420	
QY	635 TTTTACGGAATTTAAGCAGAGAAAATGATGAAGAGAAAGTCAGCTTTAATTTTCAGTAAAGG 694	
DB	421 TTTTACGGAATTTAAGCAGAGAAAATGATGAAGAGAAAGTCAGCTTTAATTTTCAGTAAAGG 480	
QY	695 AGCATGTAGCTATCCGGAGCAACATCTTCCAAAGTCAAGTACTCTGGGACCGAGTGCAT 754	
DB	481 AGCATGTAGCTATCCGGAGCAACATCTTCCAAAGTCAAGTACTCTGGGACCGAGTGCAT 540	
QY	755 GAAGCAGTAGGAAGTTCAGCTCAGTGAAGAAAGGAAAGTCTCCAGAGCTCAACTCA 814	
DB	541 GAAGCAGTAGGAAGTTCAGCTCAGTGAAGAAAGGAAAGTCTCCAGAGCTCAACTCA 600	
QY	815 GTCTAAAGAAAAGAAAAGAAAATCTGCATCTGGATGAATCATGGAGATTGAAGAGGA 874	
DB	601 GTCTAAAGAAAAGAAAAGAAAATCTGCATCTGGATGAATCATGGAGATTGAAGAGGA 660	
QY	875 AAAGAAAAGAAATGCGCCGACAGACTACTGGCTACAGCCTGAAATTAATTTGTGAAATTAAT 934	
DB	661 AAAGAAAAGAAATGCGCCGACAGACTACTGGCTACAGCCTGAAATTAATTTGTGAAATTAAT 720	
QY	935 AACCAAGAACTGGGAGAGAAATATCATTAAGAAAAGGCTATTGTTAAGGAAGTAAATCA 994	
DB	721 AACCAAGAACTGGGAGAGAAATATCATTAAGAAAAGGCTATTGTTAAGGAAGTAAATCA 780	
QY	995 CAAATATACAGCTGTTGTGAAGATGATGATTTCTGGAGACAAGCTGAAATCTGACAGAC 1054	
DB	781 CAAATATACAGCTGTTGTGAAGATGATGATTTCTGGAGACAAGCTGAAATCTGACAGAC 840	
QY	1055 TCATTTAGAGACAGTAATTTCCAGCACCCAGAGAAAAAGAAATTTCTAGTTTTAAATGGAGGCTA 1114	
DB	841 TCATTTAGAGACAGTAATTTCCAGCACCCAGAGAAAAAGAAATTTCTAGTTTTAAATGGAGGCTA 900	
QY	1115 CAGAGAAATGAGGTACCTTAGAATCCATCATGAGAGACCTTTTTCAGCTACTATCTGT 1174	
DB	901 CAGAGAAATGAGGTACCTTAGAATCCATCATGAGAGACCTTTTTCAGCTACTATCTGT 960	
QY	1175 CATTGAAACTGGCCCTTTTAAAGGAGCGCAGAGTTGAAGGAAT 1216	
DB	961 CATTGAAACTGGCCCTTTTAAAGGAGCGCAGAGTTGAAGGAAT 1002	

RESULT 3

AAx85570 standard; cDNA; 1390 BP.

XX
AC AAX85570;
XX
DT 07-SEP-1999 (first entry)
XX
DE cDNA of a gene coding for the murine kin17 protein.
XX
KW Mouse; kin17 protein; cell proliferation; fertility;
KW hyperproliferative disease; protein interaction; curved DNA;
KW HIV replication; HIV integration; repair enzyme; ss.
XX
OS Mus sp.

XX
PN FR2772046-A1.
XX
PD 11-JUN-1999.

XX
PF 09-DEC-1997; 97FR-00015536.
XX
PR 09-DEC-1997; 97FR-00015536.

XX
PA (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

XX
PI Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJP;

XX
PS WPI; 1999-359999/31.

XX
PT New DNA coding for human kin17 protein - useful for controlling cell proliferation or fertility.

XX
PT Claim 21; Page 35-36; 69pp; French.

XX
CC The present sequence encodes a murine kin17 protein. The mammalian kin17 protein is useful for preparing a medicament for controlling cell proliferation or for controlling fertility. The medicaments can also be used to treat hyperproliferative diseases. Fragments between amino acids 55 and 235 (preferably between amino acids 129 and 228) of a mammalian kin17 protein are useful for regulating the interaction between proteins and curved DNA. The fragment can be used to block replication of HIV or its integration into the human genome or to target repair enzymes to curved DNA sites. Expression vectors for kin17 can be used for controlling cell proliferation

XX
SQ Sequence 1390 BP; 482 A; 256 C; 353 G; 299 T; 0 U; 0 Other;

Query Match 64.8%; Score 990.6; DB 2; Length 1390;

Best Local Similarity 83.0%; Pred. No. 3.7e-229; Indels 6; Gaps 1;

Matches 1144; Conservative 0; Mismatches 229;

QY 66 ATGGGGAAGTCCGATTTTCTTACTCCCAAGGCTATGCCAACAGGATCAAGTCCAAAGGG 125

DB 1 ATGGGCAAGTCCGATTTTCTGAGCCCAAGGCCATGCCAATAGATTAAATCCAAAGGG 60

QY 126 CTGCAGAAAGTACGCTGGTATTTGCCAGATGTGCCAGAGCAGTGCCTGGGACGAGATGGC 185

DB 61 CTCCAGAAAGTTCGCTGGTACTGCCAGATGTGCCAAAAGCAATGCCGACGAGATGGC 120

QY 186 TTTAAGTGTATGTCATGTCGCAATCTCATCAGAGCAACTATTGCTGCTTCAGAAAAT 245

DB 121 TTTAAGTGTATGTCATGTCGCAATCTCATCAGAGCAACTATTGCTGCTTCAGAAAAT 180

QY 246 CTTACAGCACTTTATGGATTATTTTTCAGAGGAATTCGAAATGACTTTCTAGAACTTCTC 305

DB 181 CTTACAGCACTTTATGGATTATTTTTCAGAGGAATTCGAAATGACTTTCTGGAACCTTG 240

QY 306 AGGAGACGCTTTGGCACTAAAAGGGTCCACAAACATTTGTCTACAAACATACATCAGC 365

DB 241 AGGCGACGCTTTGGCACTAAAAGGGTCCACAAACATTTGTCTACAAATGAATACATCAGC 300

Qy 366 CACGAGACATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTTTACTAAG 425
Db 301 CACGAGACATCCACATGAACGCTACCCAGTGGGAGACACTGACCGACTTTACCAAG 360
Qy 426 TGGCTGGGACAGAGGCTTGTGCAGAGTGCAGCAGACACCAAAAGGCTGTATATTCAG 485
Db 361 TGGCTGGGACAGAGGCTTGTGAAGTGTGATGAGACACCAAAAGGCTGTATATTCAG 420
Qy 486 TACATAGACAGGACCCAGAACTATCCCGCGCAACTGGAACTGGAGAAAAAGAAAAAG 545
Db 421 TACATAGACAGACCCAGAACTATCCCGCGCAACTGGAACTGGAACTAGAAAAAGAGAG 480
Qy 546 CAGACCTTGATGATGAAGAAAAAATCCCAAAATTTATGAAGAGCAAGTGAAGAGGC 605
Db 481 CAAGATCTGGACGATGAAGAAAAAATCCCAAGTTTCATTTAGGAGCAGGTGAGAAGGC 540
Qy 606 CTGGAAGGGAAGAACAGGAGTCCCTACTTTTACGGAATTAAGCAGAGAAATGATGA 665
Db 541 CTGGAAGGGAAGAGCAGGACACCTGTTTTCAGAACTTTAGCCGAGAAAAATGAGNA 600
Qy 666 GAGAAAGTCAGCTTTAAATTTGAGTAAAGAGCATGTAGCTCATCCGAGCAACATCTTCC 725
Db 601 GAAAAAGTTACGTTCAATCTGAATAAAGAGCGGTGCTCAGCGGAGCTTACAACATCC 660
Qy 726 AAGTCAAGTACTCTGGGACCGAGTGCATGAAGAGCATAGGAAGTTTCAGATCAGTGAA 785
Db 661 AAGTCAAGTCTTTGGGACCAAGTGCATGAAGTGTCTGGGAGCGCAGCATCCGGGAA 720
Qy 786 CGAAAAAGATCTCCAGAGCTCAACTCAGTCTAAAGAAAGAAAGAAAAAGAAATCTGCA 845
Db 721 CGAAAAAGATCTTACAGAGCTCCCGCAGCTGC-----GAAGAAGAAAGTTCGGCC 774
Qy 846 CTGATGAATCATGGAGATTGAAGAGGAAAAAGAACTGCCCGACACAGCTACTTGG 905
Db 775 CTGATGATCATGGAGCTCGAGAGGAAAGAAAGGACCGCACCGCAGACGCTCG 834
Qy 906 CTACAGCTGAAATTTATGTGAAATTAATCAAGAAACTGGGAGAGAGAAATATCATAG 965
Db 835 TTACAGCCGGGCTGTGTGAAGATGATGACAGGTACACAGCTGTGTGAAGATGACTAC 894
Qy 966 AAAAGGCTATTGTAAAGAGTAAATTCACAAATATACAGCTGTGTGAAGATGATTGAT 1025
Db 895 AAGAAAGGCTGTGTGAAGATGATGACAGGTACACAGCTGTGTGAAGATGACTAC 954
Qy 1026 TCTGGAGACAACTGAACTTTGACCAGACTCATTTAGAGACAGATAATTCAGCACAGGA 1085
Db 955 TCTGGAGACAGCTGAACTGGACCAGACTCATTTAGAGACAGTCAATTCGCGCCCGGG 1014
Qy 1086 AAAAGAAATCTAGTTTAAATGGAGGCTACAGAGAAATGAAGTACCCTAGATTCATC 1145
Db 1015 AAAAGGGTCTAGTTTAAATGGAGGCTACAGAGAAATGAAGGACCTCTCGAATCCATC 1074
Qy 1146 AATCAGAGAGCTTTTTCAGCTACTATCGTCAATGAACTGGCCCTTTTAAAGAGCGCAGA 1205
Db 1075 AATCAGAGAGCTTTTTCAGCCACGATAGTCAATGAACTGGACCTTTTGAAGAGCGCAGA 1134
Qy 1206 GTTGAAGGAATTCATATGAAGACATTTTCTAACTTGCTGAGTTTGAAGAAATTTGTTAAC 1265
Db 1135 GTTGAAGGATTCATATGAAGACATATCTAACTTGCTGAGTTTGAAGAAATTTGATAAC 1194
Qy 1266 AATCATTAAATCTTAAAGCATCAATTTGGTGTTCGCGCAAGGCAATATGAGACTTACT 1325
Db 1195 AACATTGAAACTGTGAAGCATCAATTTGGTGTGTAGCCAAAGGCACTGTGTAACCTACT 1254
Qy 1326 GTGTTAGGGTATTTCTTTTGTATAAACAAACAGGTTTTTGAAGAAATTTACTGTATAGT 1385
Db 1255 GTGTTAGGGGATTTGTTTGTATTAATAAATAAATAATCATCTATTTAAATACTAGTAA 1314
Qy 1386 TGTTCACTAACTTTGAGAGAAATTTTAAATATGTCTCATGAGGTATCAAACTATGTAA 1444
Db 1315 TAGTTGGGTAAATTTATATAAATCTATGTTTTTTTAAAGTGTAAATAAAAAAAAAAAAA 1373

RESULT 4
AAQ79936
ID AAQ79936 standard; cDNA; 1458 BP.
XX AC AAQ79936;
XX AC AAQ79936;
DT 25-MAR-2003 (revised)
DT 06-SEP-1995 (first entry)
XX Murine Kin17 cDNA.
XX Chromosomal rearrangement; kin17 protein; SOS DNA repair system; RecA;
KW genotoxic agent; zinc finger; DNA binding protein; ss.
XX Mus musculus.
XX Key Location/Qualifiers
FT primer_bind complement(1..21)
FT /tag= b
FT /note= "Oligo L (AAQ79946) binds to complementary strand"
FT misc_feature 22..1434
FT /tag= 1
FT /label= kin17 cDNA
FT /note= "nucleotides 1-1414; the genomic DNA contains at
FT least two introns within this sequence, see Comments"
FT primer_bind 32..49
FT /tag= c
FT /note= "Oligo C (AAQ79938) binding site"
FT CDS 46..1221
FT /tag= a
FT /product= "Kin17"
FT /note= "N's in the sequence denote illegible residues"
FT primer_bind complement(67..86)
FT /tag= d
FT /note= "Oligo S (AAQ79947) binds to complementary strand"
FT primer_bind 274..297
FT /tag= e
FT /note= "Oligo D (AAQ79939) binding site"
FT primer_bind complement(339..360)
FT /tag= f
FT /note= "Oligo K (AAQ79945) binds to complementary strand"
FT primer_bind 451..474
FT /tag= g
FT /note= "Oligo J (AAQ79944) binding site"
FT primer_bind complement(550..567)
FT /tag= h
FT /note= "Oligo E (AAQ79940) binds to complementary strand"
FT primer_bind 802..825
FT /tag= i
FT /note= "Oligo F (AAQ79941) binding site"
FT primer_bind complement(839..862)
FT /tag= j
FT /note= "Oligo G (AAQ79942) binds to complementary strand"
FT primer_bind complement(1435..1458)
FT /tag= k
FT /note= "Oligo B (AAQ79937) binds to complementary strand"
XX FR2706487-A1.
PN
XX 23-DEC-1994.
PD
XX 15-JUN-1993; 93FR-00007171.
PF
XX 15-JUN-1993; 93FR-00007171.
PR
XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
PA
XX Angulo-Mora JF, Tissier A, Frelat G, Mauffrey P, Guilly M;
PI WPI; 1995-039031/06.
XX Purified murine kin17 protein prepn. for detecting chromosomal
PT

PT rearrangements - also related antibodies, human and murine DNA, primers,
 XX probes and vectors, used to assess damage caused by genotoxic agents.
 PS Claim 9; Page 33; 54pp; French.
 XX The murine Kin17 protein includes a zinc finger domain (see AAR66766),
 CC recognises single- and double-stranded DNA (partic. regions of secondary
 CC structure), has apparent mol. wt. 43 kD and is recognised by both anti-
 CC kin17 antibodies and antibodies against the RecA protein of E.coli. The
 CC kin17 protein is involved in DNA repair; it can be used to monitor
 CC chromosomal rearrangements following exposure to genotoxic agents. The
 CC kin17 cDNA sequence AA079936 consists of a 1414 nucleotide sequence,
 CC flanked by primer binding sites; the genomic kin17 DNA is claimed in
 CC which an intron is inserted at position 137 of the 1414 nucleotide cDNA
 CC and a second intron is located between nucleotides 339-429 of the cDNA.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 1458 BP; 496 A; 272 C; 372 G; 311 T; 0 U; 7 Other;
 Query Match 64.6%; Score 986.4; DB 2; Length 1458;
 Best Local Similarity 81.9%; Pred. No. 3.9e-228;
 Matches 1147; Conservative 0; Mismatches 248; Indels 6; Gaps 1;
 QY 44 AGTGATCGCTGCGTGGTCCGATCGGGAAGTCGGATTCTTCTACTCCCAAGGCTATCGC 103
 DB 24 AGTTGAGCGCTCAGGTCGCGATGGGCAANTCGGATTTCTGAGCCCGCAGCCATCGC 83
 QY 104 CAACAGGATCAAGTCCAGGGGCTCGAGAAAGTACGCTGGTATTCAGAGATGCGCAGAA 163
 DB 84 CAATAGAAATTAAGTCCAAAGGGCTCCAGAAAGCTTNNNTNNTACTGCGCAGATGCGCAAA 143
 QY 164 GAGTCCCGGACGAGAAATGGCTTTAAGTGTCAATGTATGTCGGATCTCATCAGAGACA 223
 DB 144 GCAATGCCGCGAGAAATGGCTTTAAGTGTCAATGTATGTCGGATCTCATCAGAGACA 203
 QY 224 ACTATTGCTGGCTTCAGAAATCTCCAGCAGTTTATGGAATTTTTCAGAGAAATTCGG 283
 DB 204 ACTGTTGCTGGCTTCAGAAATCTCCAGCAGTTTATGGAATTTTTCAGAGAAATTCGG 263
 QY 284 AAATGACTTTCTAGAACTTCTCAGGAGACGCTTTGGCACTAAAGGGTCCCAACAACAT 343
 DB 264 AAATGACTTTCTGGAATCTCTGAGGCGCGCTTTGGCACTAAAGGGTCCCAACAACAT 323
 QY 344 TGTCTCAACGGAATACATCAGCCACCGAGAGACATCCACATGAATGCCACTCAGTGGGA 403
 DB 324 TGTCTCAACGGAATACATCAGCCACCGAGAGACATCCACATGAATGCCACTCAGTGGGA 383
 QY 404 AACTCTGACTGATTTTACTAAGTGGCTGGCGAGAGAGGCTTGTGCAAAAGTGGACGAGAC 463
 DB 384 GACACTGACCGACTTTTACCAGTGGCTGGCGAGAGAGGCTTGTGTAAGTGGATGAGAC 443
 QY 464 ACCAAAGGCTGCTATATTACGTACATAGACAGGACCCAGAAATCTATCGCGGCAACT 523
 DB 444 ACCGAAGGCTGCTATATTACGTACATAGACAGGACCCAGAAACCTATCGTGGCAACT 503
 QY 524 GGAACTGGAGAAAGAAAGACAGACCTTGATGATGAAGAAAGAACTGCCAAATTTAT 583
 DB 504 GGAACTGGAGAAAGAAAGACAGACCTTGATGATGAAGAAAGAACTGCCAAATTTAT 563
 QY 584 TGAAGAGCAAGTCAAGAGGCTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 643
 DB 564 TGAAGAGCAAGTCAAGAGGCTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 623
 QY 644 ATTAAGCAGAGAAATGATGAAGAGAAAGTCAAGTAAATTTTGAAGTAAAGGAGCATGTAG 703
 DB 624 ACTTAGCCGAGAAATGAGGAGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 683
 QY 704 CTATCCGAGCAACATCTTTCCAAAGTCAAGTACTCTGGGACCGAGTGCATGAAGCGAT 763
 DB 684 CTGAGGGGAGTCAACATCTTCCAAAGTCAAGTACTCTGGGACCGAGTGCATGAAGCGAT 743
 QY 764 AGGAAGTTCAGCATCAGTGAAGCAAGAAAGTCTTCCAGAGCTCAACTCAGTCTAAAGA 823

DB 744 GGGGAGCGCAGCATCCGGGAAACGGAAGAGTCTTTCACAGAGCTCCGCCACGCCT----- 798
 QY 824 AAAGAGAGAAAGAAATCTGCACCTGGATGAATCATGAGAGATTCGAGAGGAGAAAGAAAG 883
 DB 799 -CGAAGAGAGAGAGAGTCCGCCCTGGATGAGATCATGAGGCTCGAAGAGGAGAAAGAAAG 857
 QY 884 AACTGCCCGAAGACAGACTACTGGCTACAGCCTCGAAATTTATTTGGAATAATTAACCAAGAA 943
 DB 858 GACCGACGGACAGAGGCTGTTTACAGCCGGGATCGTTGTGAAATTTATAACGAAGAA 917
 QY 944 ACTGGAGAGAAATATCATATAGAAAAAGGCTATTGTTAAGNAGTAAATTCACAAATATAC 1003
 DB 918 GCTTGGGAGAAATATATCAAGAAAGAAAGGGGTCGTTAAGGAAGTGTATGACAGGTACAC 977
 QY 1004 AGCTGTTCTGAGAGTATGTTGATCTTGGAGACAAAGCTGAAACTTTGACACAGACTCATTTAGA 1063
 DB 978 AGCTGTGTTAAGATGACTGACTCTGGAGACAGGCTGAAACTGGACAGACTCATTTAGA 1037
 QY 1064 GACAGTAAATTCAGCACCGAGGAAAGAAATTTCTAGTTTAAATGGAGGCTACAGAGGAAA 1123
 DB 1038 GACAGTCAATTCGGGCCCGGGGAAAGGGTCTTAGTTTAAATGGAGGCTACAGAGGAAA 1097
 QY 1124 TGAAGGTACCTTAGAATCCATCAATGAGAAAGACTTTTTCAGCTACTATCTGTCATTTGAAC 1183
 DB 1098 TGAAGGCACTCTCGAATCCATCAATGAGAAAGGCTTTTTCAGCCACGATAGTCATTTGAAC 1157
 QY 1184 TGGCCCTTTAAAGGACCGCAGAGTTGAAGGAATTCATATGAAGACATTTCTAAACTTGC 1243
 DB 1158 TGGACCTTTGAAGGACCGCAGAGTTGAAGGAATTCATATGAAGACATATCTTAACCTGC 1217
 QY 1244 CTGAGTTTGAATAATTTGTTAAACAATACATTTAAATCTTTAAAGCATCAATTCGTGTTCG 1303
 DB 1218 TTGAGTTTGAATAATTTGATACACACATTTGAACTGTGAAGCATCAATTCGTGTTCG 1277
 QY 1304 CAAGCATATGAGACTCTATCTGTTAGGGTATATCTTTTGTATATAAACAACAGGTT 1363
 DB 1278 CAAGCATGTTGTAACCTCTACTGTTAGGGGATTTGTTTGTATATAAACAACAGGTT 1337
 QY 1364 TTTGAAATATTTACTGTATAGTTGTTTCACTAAACTTTGAGAGAAATTTAAATTTATGTCCTC 1423
 DB 1338 CATCTATTAAATATCTAGTGAATAGTTGGTAAATTTATATAAATCTATGTTTATTTT 1397
 QY 1424 ATGAGGTATCAAACTATGTAA 1444
 DB 1398 AAGTGTAAAAAAGAAAAA 1418
 RESULT 5
 AAX85551
 ID AAX85551 standard; cDNA; 1002 BP.
 XX
 AC AAX85551;
 XX
 DT 07-SEP-1999 (first entry)
 XX
 DE cDNA of a gene coding for a truncated human kin17 protein.
 XX
 KW kin17 protein; cell proliferation; fertility; hyperproliferative disease;
 KW protein interaction; curved DNA; HIV replication; HIV integration;
 KW repair enzyme; ss.
 XX
 OS Homo sapiens.
 XX
 PN FR2772046-A1.
 XX
 PD 11-JUN-1999.
 XX
 PF 09-DEC-1997; 97FR-00015536.
 XX
 PR 09-DEC-1997; 97FR-00015536.
 XX
 PA (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
 XX

PI Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
XX WPI; 1999-359999/31.
DR
XX New DNA coding for human kin17 protein - useful for controlling cell
PT proliferation or fertility.
XX
XX Claim 5; Page 32; 69pp; French.
XX
CC The present sequence encodes a truncated human kin17 protein with amino
CC acids 129-228 deleted. The mammalian kin17 protein is useful for
CC preparing a medicament for controlling cell proliferation or for
CC controlling fertility. The medicaments can also be used to treat
CC hyperproliferative diseases. Fragments between amino acids 55 and 235
CC (preferably between amino acids 129 and 228) of a mammalian kin17 protein
CC are useful for regulating the interaction between proteins and curved
CC DNA. The fragment can be used to block replication of HIV or its
CC integration into the human genome or to target repair enzymes to curved
CC DNA sites. Expression vectors for kin17 can be used for controlling cell
CC proliferation
XX
SQ Sequence 1002 BP; 343 A; 189 C; 230 G; 240 T; 0 U; 0 Other;
Query Match 42.8%; Score 653.4; DB 2; Length 1002;
Best Local Similarity 76.6%; Pred. No. 1.2e-147;
Matches 969; Conservative 0; Mismatches 1; Indels 295; Gaps 2;
40 AGAAGTATCGCTGCCGTGGTGGCCATCGGGAAGTCGGATTTCTTACTCCCAAGGCTA 99
Db 32 AGAAGTATCGCTGCCGTGGTGGCCATCGGGAAGTCGGATTTCTTACTCCCAAGGCTA 91
100 TCGCCAAAGAGATCAAGTCCAAAGGGGCTGCAAGAGCTACGGTGGTATGCGAGATGTC 159
Db 92 TCGCCAAAGAGATCAAGTCCAAAGGGGCTGCAAGAGCTACGGTGGTATGCGAGATGTC 151
160 AGAGCAGTGGCGGACAGAGATGGCTTTAGTGTCAATGTCATGTCGCAATCTCATCAGA 219
Db 152 AGAGCAGTGGCGGACAGAGATGGCTTTAGTGTCAATGTCATGTCGCAATCTCATCAGA 211
220 GACAACTATTGCTGGCTTCAGAAAATCCTCAGCAGTTTATGATTTATTTTCAGAGGAAT 279
Db 212 GACAACTATTGCTGGCTTCAGAAAATCCTCAGCAGTTTATGATTTATTTTCAGAGGAAT 271
280 TCCGAAATGACTTTCTAGAACTTCTCAGAGACGCTTTGGCACTAAAGGGTCCACAACA 339
Db 272 TCCGAAATGACTTTCTAGAACTTCTCAGAGACGCTTTGGCACTAAAGGGTCCACAACA 331
340 ACATTGTCTACAAGAAATACATCAGCCACCGAGAGCAGATCCACATGAATGCCATCAGT 399
Db 332 ACATTGTCTACAAGAAATACATCAGCCACCGAGAGCAGATCCACATGAATGCCATCAGT 391
400 GGGAACTCTGACTGATTTTACTAAGTGGCTGGGAGAGAGGCTTTGTGCAAGTGGACG 459
Db 392 GGGAACTCTGACTGATTTTACTAAGTGGCTGGGAGAGAGGCTTTGTGCAAGTGGACG 444
460 AGACACCAAAAGGCTGGTATATTCAGTACATAGACAGGGAACCCAGAAACTATCCGCCGCG 519
Db 445 ----- 444
520 AACTGGAACTGGAGAAAAGAAAAGCAGGACCTTTGATGATGATGAGAAAAGAACTGCCAAAT 579
Db 445 ----- 444
580 TTAATTGAAGACAGTGAGAGAGGCTTGGAGGAGGAAACAGGAGGTCCCTACTTTTA 639
Db 445 ----- 444
640 CGGAATTAAGCAGAGAAAATGATGAAGAGAAAGTCACGTTTAAATTTGATGAAGAGCAT 699
Db 445 ----- 444
700 GTAGCTCATCGGAGCAACATCTTCCAAGTCAAGTACTCTGGGACCGAGTGCACCTGAAGA 759
|||||

Db 445 -----AGTGCATGAAGA 457
Qy 760 CGATAGGAAGTTTCAGCATCAGTGAAACGAAAGAAATCTTCCAGAGCTCAACTCAGTCTA 819
Db 458 CGATAGGAAGTTTCAGCATCAGTGAAACGAAAGAAATCTTCCAGAGCTCAACTCAGTCTA 517
Qy 820 AAGAAAAGAGAAAAGAAAATCTGCACCTGGATGAAATCATGGAGATTTGAGAGGAAAAGA 879
Db 518 AAGAAAAGAGAAAAGAAAATCTGCACCTGGATGAAATCATGGAGATTTGAGAGGAAAAGA 577
Qy 880 AAGAACTGCCCGCAACAGACTACTTGGCTCAGACCTGAAATTTATGTGAAAATTTATAACCA 939
Db 578 AAGAACTGCCCGCAACAGACTACTTGGCTCAGACCTGAAATTTATGTGAAAATTTATAACCA 637
Qy 940 AAGAACTGGGAGAGAAAATATCATAGAAAAGGCTATTGTTAAGGAAAGTAAATTCACAAAT 999
Db 638 AAGAACTGGGAGAGAAAATATCATAGAAAAGGCTATTGTTAAGGAAAGTAAATTCACAAAT 697
Qy 1000 ATACAGCTGTTGTAAGATGATTTCTGGAGACAAGCTGAAAATTTGACCGAGACTCAT 1059
Db 698 ATACAGCTGTTGTAAGATGATTTCTGGAGACAAGCTGAAAATTTGACCGAGACTCAT 757
Qy 1060 TAGAGACAGTAATTCACAGCACCGAGAAAAGAAATTTCTAGTTTTTAAATGGAGGCTACAGAG 1119
Db 758 TAGAGACAGTAATTCACAGCACCGAGAAAAGAAATTTCTAGTTTTTAAATGGAGGCTACAGAG 817
Qy 1120 GAAATGAAGGTACCTAGAAATCCATCAATGAGAAGACTTTTTTCAGCTACTATCTCATTTG 1179
Db 818 GAAATGAAGGTACCTAGAAATCCATCAATGAGAAGACTTTTTTCAGCTACTATCTCATTTG 877
Qy 1180 AAATCTGCCCTTTAAAAGGAGCGAGCTTGAAGGAATTTCAATATGAAGACATTTTCTAAAC 1239
Db 878 AAATCTGCCCTTTAAAAGGAGCGAGCTTGAAGGAATTTCAATATGAAGACATTTTCTAAAC 937
Qy 1240 TTGCTCAGTTTGAAGAAATTTGTTAAACAATA-CATTAAAATCTTAAAGCATCAAAATTTGGTG 1298
Db 938 TTGCTCAGTTTGAAGAAATTTGTTAAACAATACTTTTAAATCTTAAAGCATCAAAATTTGGTG 997
Qy 1299 TTCGC 1303
Db 998 TTCGC 1002
RESULT 6
ADQ56943
ID ADQ56943 standard; DNA; 679 BP.
XX
AC ADQ56943;
XX
DT 21-OCT-2004 (first entry)
XX
DE Novel canine microarray-related DNA sequence SeqID8245.
XX
KW canine microarray; drug screening; toxicity assay;
KW environmental pollutant; cellular response; gene expression profile;
KW toxic response; liver necrosis; fatty liver disease;
KW protein adduct formation; hepatitis; dog; ds.
XX
OS Canis familiaris.
XX
PN WO2004063324-A2.
XX
PD 29-JUL-2004.
XX
PF 05-MAY-2003; 2003WO-US013853.
XX
PR 03-MAY-2002; 2002US-0377240P.
XX
PA (GENE-) GENE LOGIC INC.
PA (PFIZ) PFIZER PROD INC.
XX
PI Diggans JC, Porter M, Wei T;
XX

DR WPI: 2004-561890/54.
XX New isolated nucleic acid molecule, useful for drug screening and
PT toxicity assays or for assessing the impact, including toxicity, of a
PT compound, pharmaceutical agent or environmental pollutant on a cell or
PT living organism.
XX Claim 1; SEQ ID NO 8245; 41pp; English.
PS This invention is related to a novel isolated canine nucleic acid
XX sequences and the construction of canine microarrays containing a
CC significant portion of the canine genome. The isolated canine nucleic
CC acid sequences of the invention may be useful for drug screening and
CC toxicity assays. The invention is therefore useful for assessing the
CC impact, including toxicity, of a compound, pharmaceutical agent or
CC environmental pollutant on a cell or living organism. The methods are
CC useful for detecting genes that are up- or down-regulated in canines in a
CC disease state. The sequences are useful as diagnostic agents or markers
CC to detect a cellular response in a sample individually or as part of a
CC gene expression profile. It is also useful as a target for agents that
CC modulate gene expression or activity. The database is useful for
CC producing electronic Northern blots that allow the user to determine the cell
CC type or tissue in which a given gene is expressed and to allow
CC determination of the abundance or expression level of a given gene in a
CC particular tissue or cell. The methods are useful for determining the
CC similarity of a toxic response to one or more individual compounds. The
CC methods are useful for predicting at least one toxic response or the
CC likelihood that a compound or test agent will induce various specific
CC pathologies such as those of the liver (liver necrosis, fatty liver
CC disease, protein duct formation or hepatitis), those of the kidney,
CC heart, brain or testes, or other pathologies associated with at least one
CC of the toxins. The methods are also useful for predicting or elucidating
CC the potential cellular pathways influenced, induced or modulated by the
CC compound or test agent due to the similarity of the expression profile
CC compared to the profile induced by a known toxin. The present sequence is
CC that of a canine DNA sequence which was claimed for use during the
CC production of a canine microarray of the invention.
XX
SQ Sequence 679 BP; 254 A; 106 C; 127 G; 185 T; 0 U; 7 Other;
Query Match 34.8%; Score 531.6; DB 13; Length 679;
Best Local Similarity 90.1%; Pred. No. 3e-118;
Matches 599; Conservative 0; Mismatches 61; Indels 5; Gaps 3;
804 AGCTCAACTCAGTCTTAAGAAAGAAAGAAAGAAATCTGCATGATCAATCATGGAG 863
1 AGCTGGCTCAGTCAAAGAAAGAAAGAAAGAAATCTGCNNCCGACGAAATCATNNAG 60
864 ATTGAAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAATCTGCATGATCAATCATGGAG 923
61 ATTGAAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAATCTGCATGATCAATCATGGAG 120
924 GTGAAATATTACCAAGAAAGAAAGAAAGAAAGAAATCTGCATGATCAATCATGGAG 983
121 GTGAAATATTACCAAGAAAGAAAGAAAGAAAGAAATCTGCATGATCAATCATGGAG 180
984 GAAGTAATTGACAAATATACAGCTGTGTGAAGATGATTTCTCGACAGCAAGCTGAAA 1043
181 GAAGTAATTGACAAATATACAGCTGTGTGAAGATGATTTCTCGACAGCAAGCTGAAA 240
1044 CTTGACAGACTCATTTAGAGACAGTAATTCAGACACGAGAAAGAAATCTGATTTTA 1103
241 CTTGACAGACTCATTTAGAGACAGTAATTCAGACACGAGAAAGAAATCTGATTTTA 300
1104 AATGGAGGCTACAGAGAAATGAAGTACCTAGATCCATCAATCAGAGACTTTTCA 1163
301 AATGGAGGCTACAGAGAAATGAAGTACCTAGATCCATCAATCAGAGACTTTTCA 360
1164 GCTACTACTGCTATTGAAACTGGCCCTTTAAAGAGCAGCAGAGTGTGAAGGAATTCATAT 1223
361 GCTACTACTGCTATTGAAACTGGCCCTTTAAAGAGCAGCAGAGTGTGAAGGAATTCATAT 420
1224 GAAGACATTTCTAACTTGCTGCTGAGTTTGAAGTTTGTGTGACATATCAATTAATCTTAA 1283

DB 421 GAAGACATTTCTAAACTTGCCTGAGTTTGAAGTCTTTTAAC-----ACATTAAGTCTTAA 477
QY 1284 AGCATCAATTTGGTGTTCGCCAAGCATTTAGAGACTCTACTGTGTAGGCTATATTTCTT 1343
DB 478 AACATCAATTTGGTGTTCGCCAAGCATTTAGAGACTCTACTGTGTAGGCTGT-TCCTT 536
QY 1344 TTGTATTAACCAACAAACAGGTTTTTGAAGTATTTACTGTATAGTTTGTTCAGCTTAACTTTGA 1403
DB 537 TTGTATTAAGCAATGAGTTTATTTAAATATTTACTGTATAGTTTGTTCAGCAAGCTTATA 596
QY 1404 GAAGTAATTATTTAGTCTCATGAGGTATCA-AACTATGTAATTTTGTCTGCTTATTTT 1462
DB 597 GCAAAATCTAATTTATGTTCTCAGGAATGATCAGAACATGTAATTTTGTTCGTTTGT 656
QY 1463 TGTCTT 1467
DB 657 TCCTT 661
RESULT 7
AAH34095
ID AAH34095 standard; cDNA; 591 BP.
XX
AC AAH34095;
XX 03-SEP-2001 (first entry)
DT Human colon cancer antigen encoding cDNA SEQ ID NO:1177.
XX
DE Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX
KW colorectal carcinoma; 88.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US026524.
XX
PR 29-SEP-1999; 99US-0157137P.
PR 03-NOV-1999; 99US-0163280P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
WPI: 2001-235357/24.
P-PSDB; AAG74690.
DR
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers.
PS Claim 1; Page 2977-2978; 9803pp; English.
XX
CC AAH32943 to AAH37195 and AAG73514 to AAG77798 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene therapy
CC and vaccine production. N and P may be used in the prevention, diagnosis
CC and treatment of diseases associated with inappropriate P expression. For
CC example, N and P may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of P by expressing inactive proteins or to
CC supplement the patient's own production of P. Additionally, N may be used
CC to produce the colon cancer-associated Ps, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the proteins. N and P
CC can be used in the prevention, diagnosis and treatment of colorectal
CC carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent
CC sequences used in the exemplification of the present invention. N.B.
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC time of publication, meaning no sequences are present for SEQ ID NO:1027

CC	to 1052, 7921 and 7922
XX	
SQ	Sequence 591 BP; 199 A; 95 C; 114 G; 182 T; 0 U; 1 Other;
Query Match	32.1%; Score 490.8; DB 4; Length 591;
Best Local Similarity	97.0%; Pred. No. 2.1e-108;
Matches 512; Conservative	0; Mismatches 12; Indels 4; Gaps 1;
Qy	1005 GCTGTGTAAGATGAATTCTTGGAGACAAGCTGAAACTTGACCGAGACTCATTTAGAG 1064
Db	
	28 GCTGCAGGAATTCGGCAGAGTCTGGAGACAAGCTGGAACCTTGACCAGACTCATTTAGAG 87
Qy	1065 ACAGTAATTCAGCACACAGGAAAAAAGAATTCCTAGTTTTAAATGAGGCTACAGAGGAAT 1124
Db	
	88 ACAGTAATTCAGCACACAGGAAAAAAGAATTCCTAGTTTTAAATGAGGCTACAGAGGAAT 147
Qy	1125 GAAGGTACCCTAGAATCCATCAATGAGAAGACTTTTTTCAGCTACTATCGTCATTTGAAACT 1184
Db	
	148 GAAGGTACCCTAGAATCCATCAATGAGAAGACTTTTTTCAGCTACTATCGTCATTTGAAACT 207
Qy	1185 GGCCCTTTAAAAGGACGACAGAGTTGAAAGGAATTCAAATATGAAGACATTTCTTAAACTTGCC 1244
Db	
	208 GGCCCTTTAAAAGGACGACAGAGTTGAAAGGAATTCAAATATGAAGACATTTCTTAAACTTGCC 267
Qy	1245 TGAGTTTGAATAATTTGTTAAACAATACATTTAAAAATCTTAAAGCATCAAATGGTGTTCCGCC 1304
Db	
	268 TGAGTTTGAATAATTTGTTAAACAATACATTTAAAAATCTTAAAGCATCAAATGGTGTTCCGCC 327
Qy	1305 AAGGCATATGAGACTCTACTGCTTAGGGTATATCTTTTGTATATAAACAAACACAGGTTT 1364
Db	
	328 AAGGCATATGAGACTCTACTGCTTAGGGTATATCTTTTGTATATAAACAAACACAGGTTT 387
Qy	1365 TTGAAAAATATTACTGTGTA----TAGTTGTTCAGCTAAACTTTTGAGAAGAAATTTAATTATGT 1420
Db	
	388 TTGAAAAATATTACTGTATAGTTAGTTGTTTCTGCTAAACTTTTGAGAAGAAATTTAATTATGT 447
Qy	1421 CTCATGAGGTATCAAACATATGTAATTTTGTCCTTTGCTTATTTTGTCTTTCTTTGTAATTTA 1480
Db	
	448 CTCATGAGGTATCAAACATATGTAATTTTGTCCTTTGCTTATTTTGTCTTTCTTTGTAATTTA 507
Qy	1481 CTTCATCAGTTTATATCTTCAATTAAGAAATGTTATTTAAAAAAAAAAAA 1528
Db	
	508 CTTCATCAGTTTATATCTTCAATTAAGAAATGTTATTTAAAAAAAAAAAA 555
RESULT 8	
AAx85550	
ID	AAx85550 standard; cDNA; 1102 BP.
AC	AAx85550;
XX	
DT	07-SEP-1999 (first entry)
XX	
DE	cDNA of a gene coding for a mouse deleted kin17 protein.
XX	
KW	kin17 protein; cell proliferation; fertility; hyperproliferative disease;
KW	protein interaction; curved DNA; HIV replication; HIV integration;
KW	repair enzyme; ss.
XX	
OS	Mus sp.
XX	
PN	FR2772046-A1.
XX	
PD	11-JUN-1999.
XX	
PF	09-DEC-1997; 97FR-00015536.
XX	
PR	09-DEC-1997; 97FR-00015536.
XX	
PA	(COMS) COMMISSARIAT ENERGIE ATOMIQUE.
XX	
PI	Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
.XX	

DR WPI; 1999-359999/31.
XX New DNA coding for human kin17 protein - useful for controlling cell
PT proliferation or fertility.
PT
XX Claim 4; Page 31; 69pp; French.
PS
XX

The present sequence encodes a mouse kin17 protein with amino acids 129-228 deleted. The mammalian kin17 protein is useful for preparing a medicament for controlling cell proliferation or for controlling fertility. The medicaments can also be used to treat hyperproliferative diseases. Fragments between amino acids 55 and 235 (preferably between amino acids 129 and 228) of a mammalian kin17 protein are useful for regulating the interaction between proteins and curved DNA. The fragment can be used to block replication of HIV or its integration into the human genome or to target repair enzymes to curved DNA sites. Expression vectors for kin17 can be used for controlling cell proliferation

XX Sequence 1102 BP; 373 A; 205 C; 272 G; 252 T; 0 U; 0 Other;
SQ

Query Match 31.1%; Score 475; DB 2; Length 1102;
 Best Local Similarity 65.4%; Pred. No. 1.8e-104;
 Matches 897; Conservative 0; Mismatches 180; Indels 294; Gaps 2

Qy 66 ATGGGGAAGTCGGATTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCCAAGGGG 125
 |||||
Db 1 ATGGCAAGTCGGATTTCTGAGCCCAAGGCCATCGCCATAGAATTAAATCCCAAGGG 60
 |||||

Qy 126 CTGCAGAAGCTACGCTGGTATGTCAGATGTCGCAAGACAGTCGCGGACGAGATCGC 185
 |||||

Db 61 CTCAGAAGCTTCGCTGGTACTGCCAGATGTGCCAAAGCAATGCGCGACGAGATGGC 120
 |||||

Qy 186 TTTAAGTGTCAATCTATGTCGGAATCTCATCAGAGACAATATTGCTGGCTTCAGAAAT 245
 |||||

Db 121 TTTAAGTGTCACTGTATGTCTGAATCTCATCAAGACAACTGTGCTGGCTTCAGAAAC 180
 |||||

Qy 246 CCTCAGCAGTTTATGGATTAATTTTCAGAGGAATTCGAAATGACTTTCTAGACTTTCTC 305
 |||||

Db 181 CCTCAGCAGTTTATGGATTAATTTTCAGAGGAAATTCGAAATGACTTTCTGAACTTCTG 240
 |||||

Qy 306 AGGAGACGCTTTGGCAGCTAAAGGGTCCACAAACATTTGCTACAAAGATACATCAGC 365
 |||||

Db 241 AGGCGACGCTTTGGCAGCTAAAGGGTCCACAAACAACTTGCTCTCAATGAATACATCAGC 300
 |||||

Qy 366 CACCGAGAGCACATCCACATGAATGCCACTCAGTGGGAAACTCTCACTGATTTTACTAAG 425
 |||||

Db 301 CACCGAGAGCACATCCACATGAACGCTACCCAGTGGGAGACACTGACCGACTTTTACCAAG 360
 |||||

Qy 426 TGGCTGGGCAGAGAGGCTTGTGCAAAAGTGGACGAGACACCAAAAGGCTGGTATATTTCAG 485
 |||||

Db 361 TGGCTGGGCAGAGAGGCTTGTGTA--
 |||||

Qy 486 TACATAGACAGGAGCCAGAAACTATCCGCGGGCACTGGAACTGGAGAAAAGAAAAG 545
 |||||

Db 388 -----
 |||||

Qy 546 CAGGACCTTGATGATGAAGAAAABACTGCCAAATTTATTGAAGAGCAAGTGAGAGAGGCG 605
 |||||

Db 388 -----
 |||||

Qy 606 CTGGAAGGGAAGGAACAGGAGGTCCTACTTTTACGGAATTAAGCAGAGAAAATGATGAA 665
 |||||

Db 388 -----
 |||||

Qy 666 GAGAAAGTCAGTTTAATTGAGTTAAAGGAGCATGTAGCTCATCCGGAGCAACATCTTCC 725
 |||||

Db 388 -----
 |||||

Qy 726 AAGTCAAGTACTCTGGGACCGAGTGCCTGAAGACGATAGGAAGTTTCAGCATCAGTGA 785
 |||||

Db 388 -----GGTACAGTGCATGAAGCTGCTGGGGAGCGCAGCATCTCCGGAAA 432
 |||||

Qy 786 CGAAAAGAAATCTTCCCAGAGCTCAACTCAGTCTTAAAGAAAAGAGAAAAGAAATCTGCA 845
 |||||

Db 433 CGGAAGAGCTCTTCAGAGAGCTCCGCCAGCT-----GCGAAGAGAGAGTCCGCC 486
 Qy 846 CTGGATGAATCATCGAGATTGAAGAGGAGAAAGAAAGAACTGCCCGACAGACTACTGG 905
 Db 487 CTGGATGAGATCATCGAGCTCGAAGAGGAGAAAGAGAGCCGACGACAGACGCTGG 546
 Qy 906 CTACAGCTGAATATTCTGGAATTAACCAAGAACTGGGAGAGAAATATCATAG 965
 Db 547 TTACAGCCGGGATCGTTGTGAATTTAATCAAGAAAGCTTTGGGAGAGAAATATCACAG 606
 Qy 966 AAAAGGCTATTGTTAAGGAAGTAATTTGACAAATATACAGCTGTTGTGAAGATGATTGAT 1025
 Db 607 AAGAAAGGGTCTGTTAAGGAAGTATTGACAGGTACACAGCTGTGTGAAGATGACTGAC 666
 Qy 1026 TCTGGAGACAACTGAACTGACAGACTCAATTTAGACAGATGATTTCCAGCACAGGA 1085
 Db 667 TCTGGAGACAGGCTGAACTGGACCAAGACTCATTTAGACAGACTCAATTCGGCCCGGGG 726
 Qy 1086 AAAGAAATCTAGTTTAAATGGAGCTACAGAGAAATGAAGTACCTTAGAATCCATC 1145
 Db 727 AAAAGGGTCTAGTTTAAATGGAGCTACAGAGAAATGAAGTACCTTAGAATCCATC 786
 Qy 1146 AATGAGAGACTTTTTCAGCTACTATCGTCAATTTGAACTGGCCCTTTAAAGAGCGCAGA 1205
 Db 787 AATGAGAGGCTTTTTCAGCCAGGATAGTCATTTGAACTGGACCTTTGAAGAGCGCAGA 846
 Qy 1206 GTTGAAGGAATCAATATGAAGACATTTCTAACTTGCCTGAGTTTGAAGAAATTTGTTAAC 1265
 Db 847 GTTGAAGGATCAATATGAAGACATCTAACTTGCCTGAGTTTGAAGAAATTTGTTAAC 906
 Qy 1266 AATACATTAATTTAAAGCATCAATTTGGTGTTCGCCAAGGCATTTAGACTCTACT 1325
 Db 907 AACACATTTGAAACTGTGAAGCATCAATTTGGTGTTCGCCAAGGCATTTAGACTCTACT 966
 Qy 1326 GTGTTAGGGTATATCTTTTGTATATAAACAAGAGGTTTGTGAAGAAATATTTAGTATAGT 1385
 Db 967 GTGTTAGGGATTTGTTTGTATATAAAGAAAGAAATCAATCTATTTAATACTAGTGAA 1026
 Qy 1386 TGTTCAGCTAAATTTTGAAGAAATTTAATTTATGCTCATGAGGTATCAAA 1436
 Db 1027 TAGTTGGGTAAATTTATATAAATCTATGTTTGTATTTTAAAGTGAAGAAAAA 1077

RESULT 9

ADP22451

ID ADP22451 standard; DNA; 1372 BP.

XX AC ADP22451;

XX 12-AUG-2004 (first entry)

XX DE Sea-squirt (Ciona intestinalis) zinc finger protein coding sequence #13.

XX DE sea-squirt; zinc finger protein; gene detection; drug development;

XX KW zinc finger protein-associated disease; gene; ds.

XX OS Ciona intestinalis.

XX XX JP2004057126-A.

XX PD 26-FEB-2004.

XX PF 31-JUL-2002; 2002JP-00222484.

XX PR 31-JUL-2002; 2002JP-00222484.

XX PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.

XX DR WPI; 2004-208711/20.

XX DR P-PSDB; ADP22452.

XX PT Novel gene encoding zinc finger protein, useful as probe in gene

detecting instruments and in development of drug for treating zinc finger protein associated diseases.

Claim 3; SEQ ID NO 25; 972pp; Japanese.

The invention comprises the amino acid and coding sequences of sea-squirt (Ciona intestinalis) zinc finger proteins. The DNA and protein sequences of the invention are useful in a gene detecting instrument. The DNA and protein sequences of the invention are useful in the development of drugs for the treatment of zinc finger protein-associated diseases. The present DNA sequence encodes a sea-squirt zinc finger protein of the invention.

Sequence 1372 BP; 500 A; 224 C; 286 G; 362 T; 0 U; 0 Other;

Query Match 24.9%; Score 380.6; DB 12; Length 1372;

Best Local Similarity 60.0%; Pred. No. 1.3e-81;

Matches 704; Conservative 0; Mismatches 439; Indels 30; Gaps 3;

Qy 69 GGGAAAGTCGGATTTTCTTACTCCCAAGGCTATCGCCAAAGGATCAAGTCCAAAGGGCTG 128

Db 52 GAGAAACCCAGGATTTCTCACTGCCCAAGGCTATCGCCAAACCGTATCAAAATCAAAAGGACTT 111

Qy 129 CAGAAAGCTACGCTGTGATTTGCCAGATGTGCCAGAGCAGTCCCGGACGAGAAATCGCTTT 188

Db 112 CAAAAGTTCCGATGTTGTTCAATGTGCCAAAACAGTGCAGAGATGAGAACCGGTTT 171

Qy 189 AAGTCTCATTTGATCTCCGAACTCTCATCAGACAACTATTCTGCTTTCAGAAATCTCTCAGG 308

Db 232 GGAGCTTTTCAGATTTCTTTTCAAGTTCCTTTTCAAAGACTTTTATGCAACTTTTGA 291

Qy 309 AGACGCTTTGGCACTTAAAGGGTCCACAAACATTTGTCTACAAAGATATACATCAGCCAC 368

Db 292 ACACGTTTCGGGACTTAAAGAGTTTCAACAATTTGTGTATACAGATATATTTCCCAT 351

Qy 369 CGAGAGCATCCACATGAATGCCACTCAGTCGGGAACTCTCAGCTGATTTTACTAAGTGG 428

Db 352 AAAGAACACATTCATATGAACGCTACAAATGGGTGACGTTAACTGCTTTCACAAATGG 411

Qy 429 CTGGGACAGAAAGGCTTTGTGCAAGTGGACGAGACACCAAAAGGCTGTATATTTCAGTAC 488

Db 412 TTGGGTAGAGAGGACACTGTAAAGTGGATCAACCGAAAGGGTTGGTTTCATACATAC 471

Qy 489 ATAGACAGGACCCAGAAACTATCCCGCGCAACTGGAACCTCGAGAACTGAGAAAGAAAGCAG 548

Db 472 ATTGATAGAGACCCATTTGGTGTGGCCACAGAAAGAACTTTGATAAACAAGCAAGG 531

Qy 549 GACCTTGATGATGAGAAAGAACTGCCAAATTTTATTGAAGACCAAGTGAAGAGAGCCCTG 608

Db 532 GAGAGAGATGATGATGAAGAGCAAGCAAAAGCAATTTGAAGAAATGTTTAAAGAGAGCAA 591

Qy 609 GAAGGAGGAAGAAACAGGAGGTCCTTACTTTTACGGAATTAAGCAGAGAAATGATGAAGAG 668

Db 592 GA-----TACTGCTAAGGATTCAGACTCCCGGAAAAATTTTACTGAG 630

Qy 669 AAAGTCACGTTTAAATTTGATGAAGAGGATGTAGCTCATCCGGAGCAACATCTTCCAAG 728

Db 631 TTGGTTCCGGAATGATGAGGAAAAAATTTGCAATAAAATCT-----GCAAAATGAATTCGAG 684

Qy 729 TCAGTACTCTGGGACCGAGTGCACCTGAAGACGATAGGAAGTTTCAGCATCAGTGAAGACGA 788

Db 685 CCAAAAGCTGCCACATCTTTCACAGCTTGGCACTTCGTGTGCTCAGCTTATCCAGTTTGGTGGC 744

Qy 789 AAAGAAATCTTCCAGAGCTCAACTCAGTCTTAAAGAAAGAAAGAAAGAAATCTGCACCTG 848

Db 745 AAGAGAAACAGCCAGCTCATCCGCTTTTAACTAGCAAGAGATCAAAAGTCTGCACCTG 804

Qy 849 GATGAATCATGGAGATTGAAGAGG---AAAAGAAAGAACTCCCGAAGACAGACTACTGG 905

Db 805 GATGAATCATTTGCAGCTGAGGAGGAAAGAAAGAGCAGAAAGTTACACACTGAACACTGG 864

PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI, 2001-656860/75.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX Claim 1; SEQ ID NO 39070; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins (ABBS7737-
CC ABBS7072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1241 BP; 371 A; 294 C; 352 G; 224 T; 0 U; 0 Other;
Query Match 19.5%; Score 298.6; DB 4; Length 1241;
Best Local Similarity 54.2%; Pred. No. 8.8e-62;
Matches 638; Conservative 0; Mismatches 524; Indels 15; Gaps 1;
64 CCATCGGGAGTGGATTTCTTACTCCCAAGGTATCGCCAAACAGGATCAAGTCCAAGG 123
67 CGATGGGTGGCGCGAGGTAGGTAGCGCCCAAGTACTCGCCCAACAGATGAATCGAAGG 126
124 GCCTCAGAGTACGCTGGTATTGCCAGATGTCGAGAGCAGTGGCGGACGAGAATG 183
127 GTCTCAGAAGTGGCTGGTACTGCCAGATGTGGAGAGCAGTGGCGGATGAAGACG 186
184 GCTTTAAGTGTCAATGTATGTCGGAATCTCATCAGAGACAATATTGCTGGCTTCAGAAA 243
187 GCTTCAAGTGCACAGGATGAGGAGTCCACAGCGCCAGTGTCTCTTTGGGACA 246
244 ATCTCTCAGCAGTTTATGGATTTATTTTCAGAGGAATTCGGAATGATCTTTAGAACTTC 303
247 ATCTGGCAAAATTCCTGCACAGTTTCAGCAAGAGTTCTCCGACGGCTACATGAGTTGC 306
304 TCAGAGAGCGTTTGGCATTAAAGGTCACACAGCAATCTCTACACGGAATACATCA 363
307 TCGCGCGCGGTTTCGACAGGAGCGCAACAGCGCCCAACAGATCTACAGGAGTACATG 366
364 GCACCGGAGCAGATCCACATGAATGTCACATGCGGAACTCTGACTGATTTTACTTA 423
367 CCACAGGAGACATCCACATGACAGCCGACCGGATGGCTCACCCTGTCGACTACGTGA 426
424 AGTGGCTGGGCGAGAGAGGCTTTGTGCMAAGTGGACGAGACACCAAGGCTGGTATATTC 483
427 AGTGGCTGGGCGGACTGGGCAAGTGTATAGCGGATGAGACGGAGAGGCTGGTTCGTCA 486
484 AGTACATAGACAGGAGCCAGAAACTATCGCGCGCACTGGAATGAGAGAAAGAGAAA 543
487 CCTACATTGATGCGAGTCCAGAGGCCATGGAACAGCAGCGGAGGCTGATCGCAAGGAGA 546
544 AGCAGGACCTTGATGATGAAGAAAAAATGCGCAAAATTTATTGAAGAGCAAGTGAAGAG 603
547 AGATGAGAAGGAGACAGAGGAGCGGATGGCGCACTTCATTGAGCAGCAGATTAATAATG 606
604 GCCTGGAAGGGAGGAGGAGGCTCCCTACTTTTACGGAATTAAGCAGAGAAAATGATG 663
607 CCAAGGCCAAGGACGGCGAGGAGGACGAGAGCCAGGAGAGGTTCAACCGAGCTAAAGCGG 666
664 AAGAGAAAGTCAGTTTAATTTAGTAAGGAGCATGTAGCTCATCGGAGCAACATCTT 723
667 AAGAGAACCAAGCACTCAAGCTTGTATTCG-----CCTTGAGAAAAAGT 711
724 CCAAGTCAAGTACTCTGGGACCGAGTGCACCTGAAGACGATAGGAAGTTCAGCATCATGTA 783

Db 712 TCCAGCCTGACACTGTGTAGGGAAATCCGCTTAGCCAAAGCAGACTGCCCCCTGAAGCTG 771
Qy 784 AACCAAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAAGAAAAGAAATCTG 843
Db 772 AAAAAAGGTGTTCAAGAAACCCAAATCCGTTGGTGGAGACAGCAACCGGTCGGTGC 831
Qy 844 CACTGGATGAATCATGGAGATTGAAGAGGAAAGAAAAGAACTGCCGGAACAGACTACT 903
Db 832 TGGACGAGATCATCAAGCAGGAGGAAAGCAAAAGAGCGGTGCCAACCGCAAGGACTACT 891
Qy 904 GGCTACAGCCCTGAATTTATTTGTAAGAAATTAACCAAGAACTGGGAGAGAAATATCATA 963
Db 892 GGCTGCACAAAGGATATCGTGGTCAAAATTTATTTCCAAATCCATGGGCGAAAAGTTCTTCA 951
Qy 964 AGAAAAGGCTATTGTTAAGGAAGTAAATGACAAATATACAGCTGTTGTTGAAGATGATTG 1023
Db 952 AAAAAAAGCGGTTGCTCGACGTAATTTGACAGATATCAGGGCAAAATCAAGTTCTTGG 1011
Qy 1024 ATTTGGAGACAGCTGAAACTTGACCAGACTTCAATTTAGACAGAGTAAATTCAGCACCAG 1083
Db 1012 AGACTGGGGAAGACTAAAGTGGATCAAGCTCATTTGGAGACGCTAATCCCGCTTTGG 1071
Qy 1084 GAAAAAGAAATTTAGTTTAAATGGAGCTACAGAGGAAATGAAAGTACCTAGAAATCCA 1143
Db 1072 ACAAGCCTGTATGGTGGTTAATGGCGCTTATCGGGGATCCGAGGCTCTGCTAAGGAAAC 1131
Qy 1144 TCAATGAGAAGACTTTTTCAGCTACTATCGTCAATTGAAACTGGGCCCTTTAAAGGACGCA 1203
Db 1132 TGGACGCGCAGATATTTCACTCAGCTGGGAAATATTCACAGGCTCTCTCAAGGCGAGAA 1191
Qy 1204 GAGTTGAAGGAATTCATATATGAAGACATTTCTAAACT 1240
Db 1192 TTGTAGACAACGTCGAATACGAAGATATATCTAAACT 1228
RESULT 12
AAAA1559
ID AAA01559 standard; cDNA; 300 BP.
XX AC AAA01559;
XX DT 19-MAY-2000 (first entry)
XX DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:1550.
XX KW Human; colon cancer; tumour; diagnosis; gene expression product; probe;
KW detection; cancerous state; metastasis; identification; breast cancer;
KW oestrogen receptor-positive breast cancer; therapy;
XX KW oestrogen receptor-negative breast cancer; lung cancer; ss.
XX OS Homo sapiens.
XX XX WO9958675-A2.
XX PN 18-NOV-1999.
XX XX 13-MAY-1999; 99WO-US010602.
XX PF 14-MAY-1998; 98US-0085426P.
XX PR 15-MAY-1998; 98US-0085537P.
XX PR 15-MAY-1998; 98US-0085696P.
XX PR 21-OCT-1998; 98US-0105234P.
XX PR 27-OCT-1998; 98US-0105877P.
XX (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Lebat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX WPI; 2000-126369/11.
DR

Db	1667	AAGAGACGAACTCAAGCTTGATATTCG-----CCTTGAGAAAAAGT	1711	PR	14-MAY-1999;	99US-0134219P.
Qy	724	CCAAGTCAAGTACTCTGGGACCGATGCACTGAAGCATAGGAAGTTCAGCATCAGTGA	783	PR	14-MAY-1999;	99US-0134221P.
Db	1712	TCCAGCTGACACTGTGCTAGGGAATCGCTCTAGCCAAGCGACCTGCCCTGAAGCTG	1771	PR	14-MAY-1999;	99US-0134370P.
Qy	784	AACGAAAGAACTTCCCAAGAGCTCACTCAGTCTAAAGAAAAGAAAGAAAAGAAATCTG	843	PR	18-MAY-1999;	99US-0134768P.
Db	1772	AAGAAAGGTGTCAAGAAACCAATTCGCTGGTGGAGACGCCAAACCGCGTCCGTGC	1831	PR	19-MAY-1999;	99US-0134941P.
Qy	844	CACTGGATGAATCATGGAGATGAAGAGGAAAAGAAAGAACTCCCGAAACAGACTACT	903	PR	20-MAY-1999;	99US-0135124P.
Db	1832	TGGACGAGATCATCAAGCAGGAGGAAAGCAAAAGGAGCGTGCCAAACCGCAAGGACTACT	1891	PR	21-MAY-1999;	99US-0135353P.
Qy	904	GGCTACAGCTGAATATTGTGAAATTTATTAACCAAGAACTGGGAGAGAAATATCAT	963	PR	24-MAY-1999;	99US-0135629P.
Db	1892	GGCTGCACAAGGGTATCGTGGTCAATTTATTTCCAAATCCATGGGCGAAAAGTTCTTCA	1951	PR	25-MAY-1999;	99US-0136021P.
Qy	964	AGAAAGGCTATTGTAAAGGAAGTAATTGCACAAATATACAGCTGTGTGAAGATGATG	1023	PR	27-MAY-1999;	99US-0136392P.
Db	1952	AACAAAGCGGTGTCTGGAGCGTAATTGACAGATATCAGGGCAAAATCAAGTTCTTGG	2011	PR	28-MAY-1999;	99US-0136782P.
Qy	1024	ATTCTGGAGCAAGCTGAACTTGACCA	1051	PR	01-JUN-1999;	99US-0137222P.
Db	2012	AGACTGGGAAAGCTAAAGTGGATCA	2039	PR	03-JUN-1999;	99US-0137502P.
RESULT 14						99US-0137724P.
AAC47248						99US-0138094P.
ID	AAC47248	standard; DNA; 1459 BP.				99US-0138540P.
XX	AC	AAC47248;				99US-0138847P.
XX	DT					99US-0139119P.
XX	DE	18-OCT-2000 (first entry)				99US-0139452P.
XX	DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 53113.				99US-0139453P.
KW	KW	Hybridisation assay; genetic mapping; gene expression control;				99US-0139457P.
KW	KW	protein identification; signal transduction pathway; metabolic pathway;				99US-0139458P.
KW	KW	promoter; termination sequence; ss.				99US-0139459P.
XX	OS	Arabidopsis thaliana.				99US-0139460P.
XX	PN	EP1033405-A2.				99US-0139461P.
XX	PD	06-SEP-2000.				99US-0139462P.
XX	PF	25-FEB-2000; 2000EP-00301439.				99US-0139463P.
XX	XX	25-FEB-1999;				99US-0139464P.
XX	XX	05-MAR-1999;				99US-0139465P.
XX	XX	09-MAR-1999;				99US-0139466P.
XX	XX	23-MAR-1999;				99US-0139467P.
XX	XX	25-MAR-1999;				99US-0139468P.
XX	XX	29-MAR-1999;				99US-0139469P.
XX	XX	01-APR-1999;				99US-0139470P.
XX	XX	06-APR-1999;				99US-0139471P.
XX	XX	08-APR-1999;				99US-0139472P.
XX	XX	16-APR-1999;				99US-0139473P.
XX	XX	19-APR-1999;				99US-0139474P.
XX	XX	21-APR-1999;				99US-0139475P.
XX	XX	23-APR-1999;				99US-0139476P.
XX	XX	28-APR-1999;				99US-0139477P.
XX	XX	30-APR-1999;				99US-0139478P.
XX	XX	30-APR-1999;				99US-0139479P.
XX	XX	04-MAY-1999;				99US-0139480P.
XX	XX	05-MAY-1999;				99US-0139481P.
XX	XX	06-MAY-1999;				99US-0139482P.
XX	XX	07-MAY-1999;				99US-0139483P.
XX	XX	11-MAY-1999;				99US-0139484P.
XX	XX	14-MAY-1999;				99US-0139485P.

PR	26-JUL-1999;	99US-0145276P.
PR	27-JUL-1999;	99US-0145913P.
PR	27-JUL-1999;	99US-0145918P.
PR	27-JUL-1999;	99US-0145919P.
PR	28-JUL-1999;	99US-0145951P.
PR	02-AUG-1999;	99US-0146386P.
PR	02-AUG-1999;	99US-0146388P.
PR	02-AUG-1999;	99US-0146389P.
PR	03-AUG-1999;	99US-0147038P.
PR	04-AUG-1999;	99US-0147204P.
PR	04-AUG-1999;	99US-0147302P.
PR	05-AUG-1999;	99US-0147192P.
PR	05-AUG-1999;	99US-0147260P.
PR	06-AUG-1999;	99US-0147303P.
PR	06-AUG-1999;	99US-0147416P.
PR	09-AUG-1999;	99US-0147493P.
PR	09-AUG-1999;	99US-0147935P.
PR	10-AUG-1999;	99US-0148171P.
PR	11-AUG-1999;	99US-0148319P.
PR	12-AUG-1999;	99US-0148341P.
PR	13-AUG-1999;	99US-0148565P.
PR	13-AUG-1999;	99US-0148684P.
PR	16-AUG-1999;	99US-0149368P.
PR	17-AUG-1999;	99US-0149175P.
PR	18-AUG-1999;	99US-0149426P.
PR	20-AUG-1999;	99US-0149722P.
PR	20-AUG-1999;	99US-0149723P.
PR	20-AUG-1999;	99US-0149929P.
PR	23-AUG-1999;	99US-0149902P.
PR	23-AUG-1999;	99US-0149930P.
PR	25-AUG-1999;	99US-0150566P.
PR	26-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151066P.
PR	27-AUG-1999;	99US-0151080P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156458P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161992P.
PR	28-OCT-1999;	99US-0161993P.
PR	29-OCT-1999;	99US-0162142P.
Query Match 14.7%; Score 224.4; DB 3; Length 1459;		
Best Local Similarity 53.6%; Pred. No. 8.4e-44;		
Matches 662; Conservative 0; Mismatches 516; Indels 58; Gaps 7;		
Qy	66	ATGGGGAAGTCGGATTTTCTTACTCCCAAGGCTATCCCAACAGGATCAAGTCCAAGGGG 125
Db	1	ATGGGTAAAGAACGATTTTCTAAACCCCAAGCGATTCGAATCGAATTAAGCCCAAGGA 60
Qy	126	CTGCAGAAAGCTACGCTGGTATTGGCAGATGTCCAGAAAGCAGTCCCGGACGAGAAATGGC 185
Db	61	CTTCAAAAGCTTCGATCGTATTGTTCAGATGTGTGAGAAACAATGCCGAGACGAGAATGGA 120
Qy	186	TTTAAAGTGTCAATGTATGTCCGAATCTCATCAGAGACAACCTATTGTCTGGCTTCAGAAAT 245
Db	121	TTCAAGTGTCAATGTATGATGAGCGAATCTCACAGAGGCGAGATGCAAGTTTTTGGGCAGAT 180
Qy	246	CCTCAGCAGTTTATGGATTAATTTTTCAGAGAAATTCGGAATAGACTTTCTAGAACTTCTC 305
Db	181	CCGACTCGTGTGTGCGATGGTTACTCTGGAAGAAATTCGAGCAGACGTTTCTCGATCTGATG 240
Qy	306	AGGAGACGCTTTGGCACTAAAGGGTCCACAACAACATTTGTTCTACACGAATACATCAGC 365
Db	241	AGCGGAGCACCCGCTTTCTCTCGTATCGCTGCCACTGTTGTTCTACAAATGATATATTAAAC 300
Qy	366	CACCGAGACACATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTTTACTTAAG 425
Db	301	GATAGACATCATGTTTCACTGAATTCGACGGAGTGGGCGAGCTTGACGGAGTTTATCAAG 360
Qy	426	TGGCTGGGACAGAGAGGCTTTGTGCAAGTGGACGAGACACCAAAAGCTGGTATATTCAAG 485
Db	361	CATCTCGGAAAACTCGTGAAGTGTGAAGTTGAAGAAACTCCTAAAGGTTGGTTTTATTACT 420
Qy	486	TACATACAGAGGACCCAGAAAACATATCCGCGGCACTGAACTGGGAGAAAAAGAAAG 545
Db	421	TATATAGATAGAGATTCAGACCTTTGTTTAAAGGAGAGGTTGAAGAAATAGAGAGTTAAG 480
Qy	546	CAGGACCTTGATGAAGAAAAAACTGCCAAATTTATTGA-----AGAGCA 592
Db	481	AGTGATTTAGCTGAGGAGGAGAGCAAGAGGGAGATTGAGACAGATTGAGAGAGCT 540
Qy	593	AGTGAGAAGAGGCGCTGGAAGGGAAGGAACAGGAGGTCCTTACTTTTACG----- 641
Db	541	GCTGAGAAATTGAATGGTGGTGGTGAAGGTGAAGGTGAACCAAGTGGTAATGATGAAGTTGTG 600
Qy	642	-----GAATTAACGAGAGAAATGATGAAGAG-AAAAGTCACG 677
Db	601	GATGATCGTGATGATGAGAGGAAGAAAGATGAAAGCTTGAGGTTGAAAGAGTGGGGTCAAG 660
Qy	678	TTTAAATTTGAGTAAAGCAGCATGTAGCTCATCCGAGCAACATCTTCCNAGTCAAGT--A 735
Db	661	GTTGGATTTGCTCTTGGTGGAGGTGTTAAACAGGTTGCTACGGGTAAAGAGAGAGGGGAA 720
Qy	736	CTCTGGGACCGAGTGCACTGAAGACGATAGGAAGTTTCAGCATCAGTGAACGAAAAAGAA 795
Db	721	AGCTCGAAACTTTTGTGGGTGAAGAGATGACAAAGTTGAGAGAGGGGAGAGAGG 780
Qy	796	CTTCCCAGAGCTCACTCAGTCTTAAGAAAGAGAAAAAGNAATCTGCATGGATGAA 855
Db	781	AAGAGAAAGTGGGAGCTCCGG--GAGGAGCGAGAGGAGAGATCAGCTTTTGGATGAGT 838
Qy	856	TCATGGAGATTGAAGAGGAAAAAGA---AAAGAACTGCCGAAACAGACTACTTGGCTACAGC 912

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 12:58:53 ; Search time 8541.29 Seconds
(without alignments)
8370.011 Million cell updates/sec

Title: US-09-555-529-1

Perfect score: 1528

Sequence: 1 ctgaattcagcgccgctg.....atgtattataaaaaaa 1528

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: gb_est1: *
2: gb_est2: *
3: gb_est3: *
4: gb_hic: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_est7: *
9: gb_gsa1: *
10: gb_gsa2: *
11: gb_gsa3: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1396	91.4	1396	4	CR595908 full-leng
2	1199.4	78.5	1339	4	CR618602 full-leng
3	1189.2	77.8	1543	4	AY609991 Sus scro
4	1065	69.7	1182	10	AY412518 Homo sapi
5	1039.8	68.0	1182	10	AY412519 Pan trogl
6	900	58.9	965	1	AL522388 AL522388
7	859.2	56.2	951	7	CO725903 ILLUMIGEN
8	849	55.6	988	7	CO726806 ILLUMIGEN
9	831	54.4	846	1	AL558811 AL558811
10	800.4	52.4	882	3	BQ221694 AGENCOURT
11	767.8	50.2	1154	10	AY412520 Mus muscu
12	757.4	49.6	1944	1	AL522387 AL522387
13	744.6	48.7	842	3	B1862461 603391378
14	744.2	48.7	936	7	CO725073 ILLUMIGEN
15	742.6	48.6	890	6	CD109094 AGENCOURT
16	741	48.5	768	7	CR986081 CR986081
17	735.8	48.2	848	1	AL558810 AL558810
18	728	47.6	748	5	BU161074 AGENCOURT
19	720.2	47.1	922	6	CD389584 AGENCOURT
20	715.4	46.8	718	8	CX866441 HBSC4_10
21	709.4	46.4	733	5	BX104466 BX104466
22	686.8	44.9	723	3	BQ186497 UI-E-EJ1-

23	683	44.7	693	7	CN256425
24	678.6	44.4	1044	7	CO724635 ILLUMIGEN
25	677.8	44.4	710	1	AM089573 x20e01.x
26	671.8	44.0	720	7	CK301009 UI-E-EJ1-
27	651.8	42.7	666	1	AL700007 DKE2p686M
28	647	42.3	783	7	CN166221 997480 MA
29	646.8	42.3	656	7	CV029155 7852 Full
30	642.8	42.1	786	1	AJ819816 AJ819816
31	642.4	42.0	701	1	AI807250 wf38f09.x
32	637	41.7	864	1	AM037867 AM037867
33	635.2	41.6	790	8	DN523891 1267468 M
34	632.4	41.4	666	3	BM739997 K-EST0010
35	630.4	41.3	1011	7	CO725332 ILLUMIGEN
36	630.2	41.2	871	1	AJ819723 AJ819723
37	627.6	41.1	682	1	AV721396 AV721396
38	627.4	41.1	638	3	BM711163 UI-E-DX1-
39	622	40.7	634	3	BM738984 K-EST0008
40	612.8	40.1	698	1	A1650375 wa90b01.x
41	608.4	39.8	942	5	BQ714508 AGENCOURT
42	602	39.4	719	7	CN792844 4127778 B
43	585.2	38.3	1077	5	BU511663 AGENCOURT
44	576.4	37.7	695	8	DN526620 1271404 M
45	576.2	37.7	605	6	CB242699 UI-CF-PNO

ALIGNMENTS

RESULT 1	CR595908	1396 bp	mRNA	linear	HTC 21-JUL-2004
CR595908	full-length cDNA clone CS0DJ015Y119 of T cells (Jurkat cell line)				
LOCUS	Cot 10-normalized of Homo sapiens (human).				
DEFINITION	CR595908.1 GI:50476715				
ACCESSION	HTC; CNSLT cdna.				
VERSION	Homo sapiens (human)				
KEYWORDS	Homo sapiens				
SOURCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.				
ORGANISM	1 (bases 1 to 1396)				
REFERENCE	Li,W.B., Gruber,C., Jessee,J. and Polayes,D.				
AUTHORS	Full-length cDNA libraries and normalization				
TITLE	Unpublished				
JOURNAL	Contact : Feng Liang Email : fliang@lifetech.com URL :				
REMARK	http://fulllength.invitrogen.com/ invitroGen Corporation 1600 Faraday Avenue				
REFERENCE	2 (bases 1 to 1396)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (20-JUL-2004) Genoscope - Centre National de Sequençage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)				
COMMENT	- Web : www.genoscope.cns.fr) 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.				
FEATURES	Location/Qualifiers				
source	1..1396				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="CS0DJ015Y119"				
	/tissue_type="T cells (Jurkat cell line) Cot 10-normalized"				
	/plasmid="pCMVSPORT_6"				
ORIGIN					
Query Match	91.4%				
Best Local Similarity	100.0%;				
Matches 1396;	Conservative				
	0; Mismatches				
	0; Indels				
	0; Gaps				

```
Qy 67 TGGGGAAGTCGGATTTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCAAAGGGCC 126
Db 1 TGGGGAAGTCGGATTTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCAAAGGGCC 60

Qy 127 TGCAGAGCTACGCTGGTATTTGCCAGATGTCGCCAGACAGTGCCTGGGACGAGATGGCT 186
Db 61 TGCAGAGCTACGCTGGTATTTGCCAGATGTCGCCAGACAGTGCCTGGGACGAGATGGCT 120

Qy 187 TTAAGTGTCAATGTATGTCGAAATCTCATCAGAGCAACTATTGCTGGCTTTCAGAAAAATC 246
Db 121 TTAAGTGTCAATGTATGTCGAAATCTCATCAGAGCAACTATTGCTGGCTTTCAGAAAAATC 180

Qy 247 CTCAGCAGTTATGGATTTATTTTCAGAGGAATTCGAAATGACTTTCTAGAACTTCTCA 306
Db 181 CTCAGCAGTTATGGATTTATTTTCAGAGGAATTCGAAATGACTTTCTAGAACTTCTCA 240

Qy 307 GGAGACGCTTTGGCACTAAAGGGTCCACACACATTTGCTACNACGAATACATCAGCC 366
Db 241 GGAGACGCTTTGGCACTAAAGGGTCCACACACATTTGCTACNACGAATACATCAGCC 300

Qy 367 ACCGAGAGCACATCCACATGAATGCCACTCAGTGGGAACTCTGACTGATTTTACTAAGT 426
Db 301 ACCGAGAGCACATCCACATGAATGCCACTCAGTGGGAACTCTGACTGATTTTACTAAGT 360

Qy 427 GCCTGGGCAGAGAAGGCTTGTGCAAAAGTGGACGAGACACAAAAGGCTGTATATTCAGT 486
Db 361 GCCTGGGCAGAGAAGGCTTGTGCAAAAGTGGACGAGACACAAAAGGCTGTATATTCAGT 420

Qy 487 ACATAGACAGGACCCAGAACTACTCCGCCGCACTCGAACTGGAGAAAAGAAAAGC 546
Db 421 ACATAGACAGGACCCAGAACTACTCCGCCGCACTCGAACTGGAGAAAAGAAAAGC 480

Qy 547 AGGA CTTTGATGATGAAGAAAAAATCGCCAAATTTATTGAAGAGCAAGTGAAGAGAGGCC 606
Db 481 AGGACCTTGATGATGAAGAAAAAATCGCCAAATTTATTGAAGAGCAAGTGAAGAGAGGCC 540

Qy 607 TGGAAAGGAAGAACAGAGAGTCCCTACTTTTACGGAATTAAGCAGAGAAAATGATGAAG 666
Db 541 TGGAAAGGAAGAACAGAGAGTCCCTACTTTTACGGAATTAAGCAGAGAAAATGATGAAG 600

Qy 667 AGAAGTCACGTTTAAATTTGATGAAGGAGCATGTAGCTCATCCGAGCAACATCTTCCA 726
Db 601 AGAAGTCACGTTTAAATTTGATGAAGGAGCATGTAGCTCATCCGAGCAACATCTTCCA 660

Qy 727 AGTCAAGTACTCTGGGACCGAGTGCATGAAGAGCATAGGAAGTTTCAGCATCAGTGAAC 786
Db 661 AGTCAAGTACTCTGGGACCGAGTGCATGAAGAGCATAGGAAGTTTCAGCATCAGTGAAC 720

Qy 787 GAAAAGAAATCTCCAGAGCTCAACTCAGTCTAAAGAAAAGAAAAGAAATCTGGAC 846
Db 721 GAAAAGAAATCTCCAGAGCTCAACTCAGTCTAAAGAAAAGAAAAGAAATCTGGAC 780

Qy 847 TGGATGAAATCATGAGATTGAAGAGGAAAAGAAAGAACTGCCGGAACAGACTACTGGC 906
Db 781 TGGATGAAATCATGAGATTGAAGAGGAAAAGAAAGAACTGCCGGAACAGACTACTGGC 840

Qy 907 TACAGCCTGAAATTTATTGTGAAAAATTATAACCAAGAAACTCGGAGAGAAAATATCATAAGA 966
Db 841 TACAGCCTGAAATTTATTGTGAAAAATTATAACCAAGAAACTCGGAGAGAAAATATCATAAGA 900

Qy 967 AAAAGGCTATTGTTAAGGAAGTAAATGACAAATATACAGCTGTTGTGGAAGATGATGATT 1026
Db 901 AAAAGGCTATTGTTAAGGAAGTAAATGACAAATATACAGCTGTTGTGGAAGATGATGATT 960

Qy 1027 CTGGAGACAAGCTCAACAGCTTCACAGACTCATTTTAGACAGAGTAATTCACGACACAGAA 1086
Db 961 CTGGAGACAAGCTCAACAGCTTCACAGACTCATTTTAGACAGAGTAATTCACGACACAGAA 1020

Qy 1087 AAAGAAATTTCTAGTTTAAATGGAGGCTACAGAGGAAATGAAGGTACCCCTAGAAATCCATCA 1146
Db 1021 AAAGAAATTTCTAGTTTAAATGGAGGCTACAGAGGAAATGAAGGTACCCCTAGAAATCCATCA 1080
```

```
Qy 1147 ATGAGAAGACTTTTTCAGCTACTATCGTCATTTGAAACTGGCCCTTTAAAAAGGACGCAGAG 1206
Db 1081 ATGAGAAGACTTTTTCAGCTACTATCGTCATTTGAAACTGGCCCTTTAAAAAGGACGCAGAG 1140

Qy 1207 TTGAAGGAATTTCAATATGAAGACATTTCTAAACTTGCCTCAGTTTGAAAAATTTGTTAAACA 1266
Db 1141 TTGAAGGAATTTCAATATGAAGACATTTCTAAACTTGCCTCAGTTTGAAAAATTTGTTAAACA 1200

Qy 1267 ATACATTAAAACTTTAAAGCATCAAAATGCTGTCGCAAGGCAATATGAGACTCTACTG 1326
Db 1201 ATACATTAAAACTTTAAAGCATCAAAATGCTGTCGCAAGGCAATATGAGACTCTACTG 1260

Qy 1327 TGTAGGGTATATCTTTTGTATAAAAACAACAGGTTTTTGAATAATTTACTGTATAGTT 1386
Db 1261 TGTAGGGTATATCTTTTGTATAAAAACAACAGGTTTTTGAATAATTTACTGTATAGTT 1320

Qy 1387 GTTCAGCTAAACTTTTGAGAAGAAATTTAATTTATGTCATCAGGATCAAACTATGTAATT 1446
Db 1321 GTTCAGCTAAACTTTTGAGAAGAAATTTAATTTATGTCATCAGGATCAAACTATGTAATT 1380

Qy 1447 TTGTCCTTGTATT 1462
Db 1381 TTGTCCTTGTATT 1396

RESULT 2
CR618602 1339 bp mRNA linear HTC 21-JUL-2004
LOCUS full-length cDNA clone CS0DB008YG18 of Neuroblastoma Cot
DEFINITION 10-normalized of Homo sapiens (human).
ACCESSION CR618602
VERSION CR618602.1 GI:50499409
KEYWORDS HTC; CNSLT_CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarctontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 1339)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
Genoscope.
REFERENCE 2 (bases 1 to 1339)
AUTHORS Direct Submission
TITLE Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
COMMENT - Web : www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
source
1. .1339
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DB008YG18"
/tissue_type="Neuroblastoma Cot 10-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN
Query Match 78.5%; Score 1199.4; DB 4; Length 1339;
Best Local Similarity 92.8%; Pred. No. 7.1e-264;
Matches 1311; Conservative 0; Mismatches 1; Indels 101; Gaps 1;
Qy 27 AGAAGTGGGTCCAGAAAGTGATCGCTGCGCTGGTCCCATGGGAAAGTCGGATTTCTT 86
Db 28 AGTACTGGGTCCAGAAAGTGATCGCTGCGCTGGTCCCATGGGAAAGTCGGATTTCTT 87
```

QY	87	ACTCCAAAGGCTATCGCCAAACAGGATCAAGTCCAGGGGCTGCAGAACTACGCTGGTAT	146
DB	88	ACTCCAAAGGCTATCGCCAAACAGGATCAAGTCCAGGGGCTGCAGAACTACGCTGGTAT	147
QY	147	TGCCAGATGTGCGCAAGCAGTCGCCGGGACGAGAAATGCGCTTTAAAGTGTCAATTTGATGTCC	206
DB	148	TGCCAGATGTGCGCAAGCAGTCGCCGGGACGAGAAATGCGCTTTAAAGTGTCAATTTGATGTCC	207
QY	207	GAATCTCATCAGAGACAACATTTGCTGCTTCCAGAAAATCCTCAGCAGTTTATGATTTAT	266
DB	208	GAATCTCATCAGAGACAACATTTGCTGCTTCCAGAAAATCCTCAGCAGTTTATGATTTAT	267
QY	267	TTTTTCAGAGGAAATCCGAAATGACTTTCTAGAACTTCTCAGGAGACGCTTTGGCACTAAA	326
DB	268	TTTTTCAGAGGAAATCCGAAATGACTTTCTAGAACTTCTCAGGAGACGCTTTGGCACTAAA	327
QY	327	AGGGTCCCAACACACATTTGCTTACAAACGAATACATCAGCCACCGAGAGACATCCACATG	386
DB	328	AGGGTCCCAACACACATTTGCTTACAAACGAATACATCAGCCACCGAGAGACATCCACATG	387
QY	387	AATGCCACTCAGTGGGAACTCTGACTGATTTTACTAAGTGTGCTGGCAGAGAAAGCTTG	446
DB	388	AATGCCACTCAGTGGGAACTCTGACTGATTTTACTAAGTGTGCTGGCAGAGAAAGCTTG	447
QY	447	TGCAAGTGGAGACACACAAAGCGTGGTATTTTCACTACATAGACAGAGGACCCAGAA	506
DB	448	TGCAAGTGGAGACACACAAAGCGTGGTATTTTCACTACATAGACAGAGGACCCAGAA	507
QY	507	ACTATCCGCGGCACTGGAATCGAGAAAAAGAAAAAGCAGGACCTTGATGATGAAGAA	566
DB	508	ACTATCCGCGGCACTGGAATCGAGAAAAAGAAAAAGCAGGACCTTGATGATGAAGAA	567
QY	567	AAAACTGCCAAATTTATTAAGAGCAGTAAGTGAAGAGGCGCTCGAAGGGAGAAACAGGAG	626
DB	568	AAAACTGCCAAATTTATTAAGAGCAGTAAGTGAAGAGGCGCTCGAAGGGAGAAACAGGAG	627
QY	627	GTCCCTACTTTTACGAAATTAAGCAGAGAAATGATGAAGAGAAAGTCAAGTTTAAATTTG	686
DB	628	GTCCCTACTTTTACGAAATTAAGCAGAGAAATGATGAAGAGAAAGTCAAGTTTAAATTTG	687
QY	687	AGTAAAGGAGCATGTAGCTCATCCGAGCAACATCTTCCAACTCAAGTACTCTGGGACCG	746
DB	688	AGTAAAGGAGCATGTAGCTCATCCGAGCAACATCTTCCAACTCAAGTACTCTGGGACCG	747
QY	747	AGTGCACTGAAGACGATAGGAAAGTTTCAGCATCAGTGAAGACGAAAGAAATCTTCCAGAGC	806
DB	748	AGTGCACTGAAGACGATAGGAAAGTTTCAGCATCAGTGAAGACGAAAGAAATCTTCCAGAGC	807
QY	807	TCAACTCAGTCTTAAGAAAAAGAAAAAGAAATCTGCACTGATGAAATCATGGAGATT	866
DB	808	TCAACTCAGTCTTAAGAAAAAGAAAAAGAAATCTGCACTGATGAAATCATGGAGATT	867
QY	867	GAAGAGAAAAAGAAAAAGAACTGCCCAACAGACTACTGCTCAGGCTGAAATTTATTTG	926
DB	868	GAAGAGAAAAAGAAAAAGAACTGCCCAACAGACTACTGCTCAGGCTGAAATTTATTTG	927
QY	927	AAAAATTATAACCAAGAACTGGGAGAGAAATATCATAGAAAAAGGCTATTGTTAAGGAA	986
DB	928	AAAAATTATAACCAAGAACTGGGAGAGAAATATCATAGAAAAAGGCTATTGTTAAGGAA	987
QY	987	GTAAATTGACAAATATACAGCTGTTGTGAAGATGATTTCTGGAGACAAGCTGAACATT	1046
DB	988	GTAAATTGACAAATATACAGCTGTTGTGAAGATGATTTCTGGAGACAAGCTGAACATT	1047
QY	1047	GACCAGACTATTTAGAGACAGTAATTTCCAGCACCAAGAAAAAGAAATTTCTAGTTTTAAAT	1106
DB	1048	GACCAGACTATTTAGAGACAGTAATTTCCAGCACCAAGAAAAAGAAATTTCTAGTTTTAAAT	1084
QY	1107	GGAGGCTTACAGAGAAATGAAGGTACCCCTAGAAATCCATCAATGAGAAGACTTTTTCAGCT	1166
DB	1085	-----	1084
QY	1167	ACTATCGTCAATGAACCTGCCCCCTTTAAAAAGGACGACAGAGTTGAAGAAATTCATATATGAA	1226
DB	1085	-----GGCCCCCTTTAAAAAGGACGACAGAGTTGAAGAAATTCATATATGAA	1126
QY	1227	GACATTTCTAAACCTTGCTGAGTTTGAATTTTGAACATACATTAACATCCTTAAAGC	1286
DB	1127	GACATTTCTAAACCTTGCTGAGTTTGAATTTTGAACATACATTAACATCCTTAAAGC	1186
QY	1287	ATCAAAATTTGGTGTTCGCAAGGCATTAATGAGACTCTACTGTGTAGGATATATCTTTT	1346
DB	1187	ATCAAAATTTGGTGTTCGCAAGGCATTAATGAGACTCTACTGTGTAGGATATATCTTTT	1246
QY	1347	TATAAACAAACACAGGTTTTTGAATAATTAATCTGTATATAGTTTTCAGCTAAACCTTTGAGAA	1406
DB	1247	TATAAACAAACACAGGTTTTTGAATAATTAATCTGTATATAGTTTTCAGCTAAACCTTTGAGAA	1306
QY	1407	GAATTTAATTAATCTCTCATGAGGTATCAAACTA	1439
DB	1307	GAATTTAATTAATCTCTCATGAGGTATCAAACTA	1339
RESULT 3			
AY609991 1543 bp mRNA linear HTC 31-JAN-2005			
SUBSCROFA clone Clu_5302.scr.msx.pl.Contig4, mRNA sequence.			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
misc_feature			
ORIGIN			
Query Match 77.8%; Score 1189.2; DB 4; Length 1543;			
Best Local Similarity 89.3%; Pred. No. 1.6e-261;			
Matches 1314; Conservative 0; Mismatches 152; Indels 5; Gaps 3;			
QY	55	CCGTGCTGCCATGGGAAGTCGGATTTTCTTACTCCCAAGGCTATCGCCAAACAGGATCA	114
DB	25	CTGTGCTCCGATGGGAAGTCGGATTTTCTTACCCGNAAGGCCATCGCCACAGGATCA	84
QY	115	AGTCCAAGGGGCTCAGAGCTACCGTGGTATTTGCCAGATGTGCCAGAGCAGTGCCTGG	174
DB	85	ATCCAAGGGGCTTCCAGAGCTGCGCTGGTATTTGCCAGATGTGCCAGAGCAGTGCCTGG	144
QY	175	ACGAGAAATGGCTTTAAGTGTCTATGTCGGAATCTCATCAGAGACAACTATTTCCTGG	234
DB	145	ACGAGAAATGGCTTTAAGTGTCTATGTCGGAATCTCATCAGAGACAACTATTTCCTGG	204

[illegible]

RESULT 5
AY412519
LOCUS

AY412519	1182 bp	DNA	linear	CSS 16-DEC-2003
-				

DEFINITION	Pan troglodytes KIN gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.
ACCESSION	AY412519
VERSION	AY412519.1 GI:39768484
KEYWORDS	GSS.
SOURCE	Pan troglodytes (chimpanzee)
ORGANISM	Pan troglodytes
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Pan.
AUTHORS	1 (bases 1 to 1182) Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Perrier,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
TITLE	Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios
JOURNAL	Science 302 (5652), 1960-1963 (2003)
PUBMED	14671302
AUTHORS	2 (bases 1 to 1182) Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Perrier,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
TITLE	Direct Submission
JOURNAL	Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
COMMENT	This sequence was made by sequencing genomic exons and ordering them based on alignment.
FEATURES	Location/Qualifiers
source	1..1182 /organism="Pan troglodytes" /mol_type="genomic DNA" /db_xref="taxon:9598" <1..>1182 /gene="KIN" /locus_tag="HCM4584"
gene	
ORIGIN	
Query Match	68.0%; Score 1039.8; DB 10; Length 1182;
Best Local Similarity	88.1%; Pred. No. 2.7e-227;
Matches 1041, Conservative	0; Mismatches 141; Indels 0; Gaps
Qy	66 ATGGGGAAGTCGAGATTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCCAAGGGG 12
Db	1 ATGGGGAAGTCGAGATTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCCAAGGGG 12
Qy	126 CTCGAGAGCTACGCTGTTATTTGCCAGATGTGCCAAGCAGTCCGGGACGAGAATGCG 18
Db	61 CTGCAAGAGCTACGCTGTTATTTGCCAGATGTGCCAAGCAGTCCGGGACGAGNNNNNN 12
Qy	186 TTTAAGTGTCAATTGTATGTCGGAATCTCATCGAGACAACTATTGCTGGCTTCAGAAAAAT 24
Db	121 NNN 18
Qy	246 CCTCAGCAGTTTATGGATTAATTTTCAGAGGAATTCGGAATGCACCTTCTAGACTTCTC 30
Db	181 NNN 24
Qy	306 AGGAGACGCTTTGGCACTAAAGGGTCCACAAACAACTGTCTACAAACGAATACATCAGC 36
Db	241 NNN 30
Qy	366 CACCGAGAGCAGCATCCACATGAATGGCACTCAGTGGGAAAACCTGACTGATTTTACTAAG 42
Db	301 CACCGAGAGCAGCATCCACATGAATGGCACTCAGTGGGAAAACCTGACTGATTTTACTAAG 36
Qy	426 TGGCTGGGCGAGAGAGGCTTGTGCAAGTGGAGCGAGACACCAAAAGCGCTGGTATATTCAG 48
Db	361 TGGCTGGGCGAGAGAGGCTTGTGCAAGTGGAGCGAGACACCAAAAGCGCTGGTATATTCAG 42
Qy	486 TACATAGACAGGGACCCAGAAACTATCCCGCCGCAACTGGAACTGGAGAAAAAGAAAAAG 54

Db 421 TACATAGACAGGCCAGAACTATCCCGCGCACTGGAAGTGGAGAGAAAGAAAG 480
Qy 546 CAGGACCTTGATGATGAGAGAAAACTGCGCAATTTATTGAAGAGCAAGTGAAGAGGC 605
Db 481 CAGGACCTTGATGATGAGAGAAAACTGCGCAATTTATTGAAGAGCAAGTGAAGAGGC 540
Qy 606 CTGGAAGGAGGAGACAGAGGTCCTACTTTTTCGGAATTAAGCAGAGAAATGATGA 665
Db 541 CTGGAAGGAGGAGACAGAGGTCCTACTTTTTCGGAATTAAGCAGAGAAATGATGA 600
Qy 666 GAGAAAGTCACGTTTAAATTTAGTAAGAGCATGTAGCTCATCCGGAGCAACATCTTCC 725
Db 601 GAGAAAGTCACGTTTAAATTTAGTAAGAGCATGTAGCTCATCCGGAGCAACATCTTCC 660
Qy 726 AAGTCAAGTACTCTCGGACCGAGTGCATGAAGACGATAGGAAGTTCAGCATCAGTGAAA 785
Db 661 AAGTCAAGTACTCTCGGACCGAGTGCATGAAGACGATAGGAAGTTCAGCATCAGTGAAA 720
Qy 786 CGAAAGAAATCTTCCAGAGCTCACTCAGTCTAAGAAAGAAAGAAAGAAATCTGCA 845
Db 721 CGAAAGAAATCTTCCAGAGCTCACTCAGTCTAAGAAAGAAAGAAAGAAATCTGCA 780
Qy 846 CTGGATGAATCATGGAGATTGAAGAGAGAAAGAAAGAACTGCCGACACAGACTACTGG 905
Db 781 CTGGATGAATCATGGAGATTGAAGAGAGAAAGAAAGAACTGCCGACACAGACTACTGG 840
Qy 906 CTACAGCTGAAATTTATTTGTGAAAATTATAACCAAGAACTGGGAGAGAAATATCATAG 965
Db 841 CTACAGCTGAAATTTATTTGTGAAAATTATAACCAAGAACTGGGAGAGAAATATCATAG 900
Qy 966 AAAAGGCTATTGTTAAGAGATTAATTCACAATATACAGCTGTTGTGAAGATGATTAT 1025
Db 901 AAAAGGCTATTGTTAAGAGATTAATTCACAATATACAGCTGTTGTGAAGATGATTAT 960
Qy 1026 TCTGAGAGCAAGCTGAAACTTGACAGACTCATTTAGAGACAGTAATTCAGCACCAGGA 1085
Db 961 TCTGAGAGCAAGCTGAAACTTGACAGACTCATTTAGAGACAGTAATTCAGCACCAGGA 1020
Qy 1086 AAAAGAAATCTAGTTTTAAATGGAGGCTACAGAGAAATGAAGTACCCTAGAATCCATC 1145
Db 1021 AAAAGAAATCTAGTTTTAAATGGAGGCTACAGAGAAATGAAGTACCCTAGAATCCATC 1080
Qy 1146 AATGAGAGACTTTTTCAGCTACTATCTGATTAAGAACTGGCCCTTAAAGAGCGCAGA 1205
Db 1081 AATGAGAGACTTTTTCAGCTACTATCTGATTAAGAACTGGCCCTTAAAGAGCGCAGA 1140
Qy 1206 GTTGAAGAAATCAATATGAAGACATTTCTAAACTTTGCTCTGA 1247
Db 1141 GTTGAAGAAATCAATATGAAGACATTTCTAAACTTTGCTCTGA 1182

RESULT 6
LOCUS AL522388
DEFINITION AL522388 Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED Homo sapiens
cDNA clone CS0DB008Y18 5-PRIME, mRNA sequence.
ACCESSION AL522388
VERSION AL522388.3 GI:45697719
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 965)
Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On Feb 13, 2001 this sequence version replaced gi:31040660.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 5543.r

For more information about this cluster, see
http://www.genoscope.cns.fr/cdna/?s=CS0DB008BD09QP1&c=5543.r.
Location/Qualifiers
1. 965
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DB008Y18"
/tissue_type="NEUROBLASTOMA COT 10-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match 58.9%; Score 900; DB 1; Length 965;
Best Local Similarity 99.5%; Pred. No. 2.9e-195;
Matches 933; Conservative 1; Mismatches 1; Indels 3; Gaps 3;
Qy 28 GAACCTGGGTCAGAAAGTATGATCCTCCGCTGGTGGCCATGGGGAAGTCTGATTTCTTA 87
Db 29 GATCTGGGTCAGAAAGTATGATCCTCCGCTGGTGGCCATGGGGAAGTCTGATTTCTTA 88
Qy 88 CTCCTCAAGGCTATGCCCAACAGGATCAAGTCCCAAGGGGCTGCAGAGCTAGCGTGTATT 147
Db 89 CTCCTCAAGGCTATGCCCAACAGGATCAAGTCCCAAGGGGCTGCAGAGCTAGCGTGTATT 148
Qy 148 GCCAGATGTGCGAGAGCAGTGGCGGACGAGAAATGCTTTAAAGTGTCTATGTATGTCG 207
Db 149 GCCAGATGTGCGAGAGCAGTGGCGGACGAGAAATGCTTTAAAGTGTCTATGTATGTCG 208
Qy 208 AATCTCATCAGAGACAACTATTGCTGGCTTCAGAAAATCCTCAGCAGTTTATGATTTAT 267
Db 209 AATCTCATCAGAGACAACTATTGCTGGCTTCAGAAAATCCTCAGCAGTTTATGATTTAT 268
Qy 268 TTTTCAGAGGAAATCCGAAATCAGCTTTCTAGAACTTTCTCAGAGAGCGCTTTGGCACTAAA 327
Db 269 TTTTCAGAGGAAATCCGAAATCAGCTTTCTAGAACTTTCTCAGAGAGCGCTTTGGCACTAAA 328
Qy 328 GGGTCCCAACAACTATTGCTGTACAAAGAAATACATCAGCCACCGAGAGCACAATCCACATGA 387
Db 329 GGGTCCCAACAACTATTGCTGTACAAAGAAATACATCAGCCACCGAGAGCACAATCCACATGA 388
Qy 388 ATGCCACTCAGTGGGAACTCTGACTGATTTTACTAAGTGGCTGGGAGAGAGCGTTGT 447
Db 389 ATGCCACTCAGTGGGAACTCTGACTGATTTTACTAAGTGGCTGGGAGAGAGCGTTGT 448
Qy 448 GCAAGTGGAGAGACACCAAAAGGCTGTATTTTCTAGTACATAGACAGGACCCAGGAA 507
Db 449 GCAAGTGGAGAGACACCAAAAGGCTGTATTTTCTAGTACATAGACAGGACCCAGGAA 508
Qy 508 CTATCCCGCGCAACTGGAACTGGAGAAAAAGAAAAAGAGCAGGACCTTGTATGATGAAGAA 567
Db 509 CTATCCCGCGCAACTGGAACTGGAGAAAAAGAAAAAGAGCAGGACCTTGTATGATGAGNAA 568
Qy 568 AAACCTGCCAAATTTATTGAAGAGCAAGTGAAGAGCGCTGGAAAGGAGAAAGCAAGGAG 627
Db 569 AAACCTGCCAAATTTATTGAAGAGCAAGTGAAGAGCGCTGGAAAGGAGAAAGCAAGGAG 628
Qy 628 TCCCTACTTTTACGGAATTAAGCAGAAAAATGATGAAGAGAAATCAGCTTTAATTTGA 687
Db 629 TCCCTACTTTTACGGAATTAAGCAGAAAAATGATGAAGAGAAATCAGCTTTAATTTGA 688
Qy 688 GTAAAGGAGCATGTAGCTCATCCGGAGCAACATCTTCCAAAGTCAAGTACTCTGGGACCGA 747
Db 689 GTAAAGGAGCATGTAGCTCATCCGGAGCAACATCTTCCAAAGTCAAGTACTCTGGGACCGA 748

FEATURES
source


```
QY 748 GTGCACTGAAGAGGATAGCAAGTTTCAGCATCAGTGAACGAAAGAAATCTTCCAGAGCT 807
Db 749 GTGCACTGAAGAGGATAGCAAGTTTCAGCATCAGTGAACGAAAGAAATCTTCCAGAGCT 807
QY 808 CAACTCAGTCTTAAGAAAAAGAAAGAAAG- AAATCTGCACCTGGATGAATCATGAGAGATT 866
Db 808 CAACTCAGTCTTAAGAAAAAGAAAGAAAGAAATCTGCACCTGGATGAATCATGAGAGATT 867
QY 867 GAAGAGG-AAAAGAAAGAAAGTCCCGAAGAGAGTCTGCTACAGCTGAAATATTGT 925
Db 868 GAAGAGGAAAAAGAAAGAAAGTCCCGAAGAGAGTCTGCTACAGCTGAAATATTGT 927
QY 926 GAAATATTATACCAAGAACTGGAGAGAAATATCATA 963
Db 928 GAAATATTATACCAAGAACTGGAGAGAAATATCATA 965

RESULT 7
LOCUS CO725903 951 bp mRNA linear EST 27-JUL-2004
DEFINITION ILLUMIGEN MCQ_40291 Katze_MMPB2 Macaca mulatta cDNA clone
IBIUM:27173 5' similar to Bases 5 to 950 highly similar to human
KIN (Hs.123647), mRNA sequence.
ACCESSION CO725903
VERSION 1 GI:50707056
KEYWORDS EST.
SOURCE Macaca mulatta (rhesus monkey)
ORGANISM Macaca mulatta
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Cercopithecoidea; Cercopithecinae; Macaca.
1 (bases 1 to 951)
Magness,C.L., Fellin,P.C., Thomas,M.J., Korth,M.J., Agy,M.B.,
Prolli,S.C., Fitzgibbon,M., Scherer,C.A., Miner,D.G., Katze,M.G. and
Iadonato,S.P.
Analysis of the Macaca mulatta transcriptome and the sequence
divergence between Macaca and human
Genome Biol. 6 (7), R60 (2005)
15998449
Contact: C. Magness
Illumigen Biosciences Inc.
2203 Airport Way S, Suite 450, Seattle, WA 98134, USA
Tel: 2063780400
Fax: 2063780408
Email: cmagness@illumigen.com
Sequenced on 2004.05.28. 723 Q20 bases. Library Preparation: Prof.
Michael Katze Lab at University of Washington DNA Sequencing:
Illumigen Biosciences Inc. For further information, see
http://www.macaque.org
PCR Primers
FORWARD: CCCTCACTAAAGGGAAACAAA
BACKWARD: CACTATAGGGCGAATTGGGTA
Insert Length: 951 Std Error: 0.00
Plate: CL000321 row: C column: 12
Seq primer: CCCTCACTAAAGGGAAACAAA
POLYA=No.

FEATURES             Location/Qualifiers
     1..951
         /organism="Macaca mulatta"
         /mol_type="mRNA"
         /strain="Indian"
         /db_xref="taxon:9544"
         /clone="IBIUM:27173"
         /sex="male"
         /tissue_type="blood"
         /cell_type="PBMC"
         /dev_stage="adult"
         /lab_host="Electromax DH10B"
         /clone_lib="Katze MMPB2"
         /note="Vector; pDONR 222; Site 1: BsrG I; Site 2: BsrG I;
Created from CloneMiner cDNA Library Construction kit
(catalog #18249-029)"
```

ORIGIN

```
Query Match      56.2%; Score 859.2; DB 7; Length 951;
Best Local Similarity 96.3%; Pred. No. 6.4e-186;
Matches 901; Conservative 0; Mismatches 33; Indels 2; Gaps 2;

QY 28 GAACCTGGGGTCCAGAAAGTGATCGCTGCGGTGCGCCATGGGGAAGTCGGATTTCTTTA 87
Db 15 GCACCTAGGGTCCAGAAAGTGACCGCTGCGGTGCGCCATGGGGAAGTCGGATTTCTTTA 74
QY 88 CTCCTCAAGGCTATCCCAACAGGATCAAGTCCAAAGGGCTGCAGAAAGTACCTCGTATT 147
Db 75 CTCCTCAAGGCTATCCCAACAGGATCAAGTCCAAAGGGCTGCAGAAAGTACCTCGTATT 134
QY 148 GCCAGATGTGCCAGAAAGCAGTGCCTGGGACGAGAAATGGCTTTAAAGTGTCTATTGTTCG 207
Db 135 GCCAGATGTGCCAGAAAGCAGTGCCTGGGACGAGAAATGGCTTTAAAGTGTCTATTGTTCG 194
QY 208 AATCTCATCAGAGACAACTATTGTGCTTTCAGAAAAATCTCTCAGCAGTTTATGGATTATT 267
Db 195 AATCTCATCAGAGACAACTATTGTGCTTTCAGAAAAATCTCTCAGCAGTTTATGGATTATT 254
QY 268 TTTTCAGAGGAAATTCGAAATGACTTTCTAGAACTTCTCAGGAGACGCTTTGGCAGCTAAAA 327
Db 255 TTTTCAGAGGAAATTCGAAATGACTTTTCTAGAACTTCTCAGGAGACGCTTTGGCAGCTAAAA 314
QY 328 GGGTCCCAACAACTATTGTCTACAAAGAAATACATCAGCCAGAGACATCCACATGA 387
Db 315 GGGTCCCAACAACTATTGTCTACAAAGAAATGATAGTACATCAGCCAGAGACATCCATATGA 374
QY 388 ATGCCACTCAGTGGGAAACTCTGACTGATTTTAACTAAGTGGCTGGGACAGAGAGGTTGT 447
Db 375 ATGCCAACTCAGTGGGAAACTCTGACCGATTTTAACTAAGTGGCTGGGACAGAGAGGTTGT 434
QY 448 GCAAGTGTGACAGAGACACCAAAAGGCTGTATATTCAGTACATAGACAGGACCCAGAAA 507
Db 435 GCAAAAGTGTGAGAGACACCAAAAGGCTGTGTATATTCAGTACATAGACAGGACCCAGAAA 494
QY 508 CTATCCGCGGCAACTTGGAACTGGAGAAAAAGAAAAAGCAGGACCTTGATGATGAAGAAA 567
Db 495 CTATTCGCGGCAACTTGGAACTGGAGAAAAAGAAAAAGCAGGACCTTGATGATGAAGAAA 554
QY 568 AAACCTGCCAAATTTATTGAAGACAAAGTGAAGAGGCTTGGAAAGGAAAGGAAACAGGAGG 627
Db 555 AAACCTGCCAAATTTATTGAAGAAACAAAGTGAAGAGGCTTGGAAAGGAAAGGAAACAGGAGG 614
QY 628 TCCTACTTTTACGGAAATTAAGCAGAGAAATATGATGAAGAAAGTCAAGTAAATTTGA 687
Db 615 CCCTACTTTTACGGAAATTAAGCAGAGAAATATGATGAAGAAAGTCAAGTAAATTTGA 674
QY 688 GTAAAGGAGCATGTAGCTCATCCGAGCAACATCTTCCAAAGTCAAGTACTCTGGGACCGA 747
Db 675 GTAAAGGAGCATGTAGCTCATCCGAGCAACGCTCTTCCAAAGTCAAGTACTCTGGGACCGA 734
QY 748 GTGCACTGAAGACGATAGGAAGTTTCAGCATCAGTGAAGAAAGAAAGAAATTTTCCAGAGCT 807
Db 735 GTGCACTGAAGACAAATAGGAGTTTCAGCATCAGTGAAGAAAGAAAGAAATTTTCCAGAGCT 794
QY 808 CAACTCAGTCTAAAGAAAAAGAAAGAAAGAAATCTCAGCTGGATGAATCATCGAGAA- TT 866
Db 795 CACCTCAGTCTCAAAAAAGAAAGAAAGAAAGAAATCTCAGCTGGATGAATCATCGAGAAATTT 854
QY 867 GAAGAGGAAAGAAAGAAAGAAAGTCCCGAAGAGAGTCTGCTCAGAGCTGAAATTT-ATTGT 925
Db 855 GAAAGAGGAAAGAAAGAAAGAAAGTCCCGAAGAGAGTCTGCTCAGAGCTGAAATTTATTGT 914
QY 926 GAAATATTATACCAAGAAAGTGGAGAGAAATATCA 961
Db 915 GAAATATTATACCAAGAAAGTGGAGAGAAATATCA 950

RESULT 8
CO726806
```

LOCUS C0726806 988 bp mRNA linear EST 27-JUL-2004
DEFINITION ILLUMIGEN_MCO_48930 Katze_MMD Macaca mulatta cDNA clone
IBUW:27517 5'-similar to Bases 5 to 967 highly similar to human
KIN (Hs.123647), mRNA sequence.

ACCESSION C0726806
VERSION
KEYWORDS
SOURCE EST.
ORGANISM Macaca mulatta (rhesus monkey)
Macaca mulatta
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Cercopithecoidea; Cercopithecinae; Macaca.

REFERENCE 1 (bases 1 to 988)
AUTHORS Magness,C.L., Fellin,P.C., Thomas,M.J., Korth,M.J., Agy,M.B.,
Prohl,S.C., Fitzgibbon,M., Scherer,C.A., Miner,D.G., Katze,M.G. and
Iadonato,S.P.
TITLE Analysis of the Macaca mulatta transcriptome and the sequence
divergence between Macaca and human
JOURNAL Genome Biol. 6 (7), R60 (2005)
PUBMED 15998449
COMMENT Contact: C. Magness
Illumigen Biosciences Inc.
2203 Airport Way S, Suite 450, Seattle, WA 98134, USA
Tel: 2063780400
Fax: 2063780408
Email: cmagness@illumigen.com
Sequenced on 2004.07.02. 706 Q20 bases. Library Preparation: Prof.
Michael Katze Lab at University of Washington DNA Sequencing:
Illumigen Biosciences Inc. For further information, see
http://www.macaque.org
PCR Primers
FORWARD: CCTCCTAAAGGGACAAAA
BACKWARD: CACTATAGGCGAATGGTA
Insert Length: 988 Std Error: 0.00
Plate: CL000405 row: C column: 07
Seq primer: CCTCCTAAAGGGACAAAA
POLYA=Yes

FEATURES
source location/Qualifiers
1..988
/organism="Macaca mulatta"
/mol_type="mRNA"
/strain="Indian"
/db_xref="taxon:9544"
/clone="IBUW:27517"
/sex="male"
/dev_stages="adult"
/lab_host="Electronmax DH10B"
/clone_lib="Katze_MMD"
/note="Organ: duodenum; Vector: pDONR 222; Site 1: BsrG I;
Site 2: BsrG I; Created from Cloneminer cDNA Library
Construction kit (catalog #18249-029)"

ORIGIN

Query Match 55.6%; Score 849; DB 7; Length 988;
Best Local Similarity 94.0%; Pred. No. 1.4e-183;
Matches 915; Conservative 0; Mismatches 55; Indels 3; Gaps 3;

QY 39 CAGAAAGTGATCGCTCCGCTGTCGCCATGGGGAAGTCGGATTCTTACTCCCAAGGCT 98
Db |||||
5 CAGAAAGTGACCGCTGCTGCTGTCGCCATGGGGAAGTCGGATTCTTACTCCCAAGGCT 64
QY 99 ATCGGCAACAGAGATCAAGTCCAAAGGGCTGCGAGAGCTACGCTGGTATGCCAGATGTGC 158
Db |||||
65 ATCGGCAACAGAGATCAAGTCCAAAGGGCTGCGAGAGCTACGCTGGTATGCCAGATGTGC 124
QY 159 CAGAAAGTGTGCGGAGCAGAGATGGCTTTAAGTGTGATTGTATGCGAATCTCATCAG 218
Db |||||
125 CAGAAAGTGTGCGGAGCAGAGATGGCTTTAAGTGTGATTGTATGCGAATCTCATCAG 184
QY 219 AGACAACTATTGCTGGCTTCAGAAAATCTCAGCAGTTTATGGATTATTTTTCAGAGGAA 278
Db |||||
185 AGACAACTATTGCTGGCTTCAGAAAATCTCAGCAGTTTATGGATTATTTTTCAGAGGAA 244

QY 279 TTCGAAATGATTTCTAGAACTTTCTCAGGAGAGCGTTTGGCACTAAAGGGTCCACAC 338
Db |||||
245 TTCGAAATGATTTCTAGAACTTTCTCAGGAGAGCGTTTGGCACTAAAGGGTCCACAC 304
QY 339 AACATTGTCTACAAAGCAATACATCAGCCACCAGAGAGCACATCCACATGAATGCCACTCAG 398
Db |||||
305 AACATTGTCTACATGAGTACATCAGCCACCAGAGAGCACATCCATATGAATGCAACTCAG 364
QY 399 TGGAAACTCTGACTGATTTTAAAGTGGCTGGCGAGAGAGCGTTGTGCAAAAGTGAC 458
Db |||||
365 TGGAAACTCTGACCGATTTTAAAGTGGCTGGCGAGAGAGCGTTGTGCAAAAGTGAC 424
QY 459 GAGACACCAAAAGCGCTGATATTTACGTACATACAGAGGACCCAGAAAATCTATCGCCGG 518
Db |||||
425 GAGACACCAAAAGCGCTGATATTTACGTACATACAGAGGACCCAGAAAATCTATCGCCGG 484
QY 519 CAACCTGGAACCTGGAGAAAAAGAAAAAGCAGGACCTTGATGATGAAGAAAAAACTGCCAAA 578
Db |||||
485 CAACCTGGAACCTGGAGAAAAAGAAAAAGCAGGACCTTGATGATGAAGAAAAAACTGCCAAA 544
QY 579 TTTATTGAAGCAAGTGAGAAAGCGCTTGAAGGAAAGGAAACAGGAGGTCCCTACTTTT 638
Db |||||
545 TTTATTGAAGCAAGTGAGAAAGCGCTTGAAGGAAAGGAAACAGGAGGTCCCTACTTTT 604
QY 639 ACGGAATTAAGCAGAGAAAAATGATGAAGAAAGCTCAGTTTAATTTGAGTAAAGGAGCA 698
Db |||||
605 ACGGAATTAAGCAGAGAAAAATGATGAAGAAAGCTCAGTTTAATTTGAGTAAAGGAGCA 664
QY 699 TGTAAGTCTATCCGGAGCAACATCTTCCAAGTCAAGTACTCTGGGACCGAGTGCATGAAG 758
Db |||||
665 TGCAGTTTATCCGGAGCAACGCTTCCAAAGTCAAGTACTCTGGGACCGAGTGCATGAAG 724
QY 759 ACGATAGAAAGTTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 818
Db |||||
725 ACAATAGAAAGTTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 784
QY 819 AAGAAAGAGAAAGAAAGAAATCTGCACTGGATCAATCATGGAGATTGGAAGAGAAAG 878
Db |||||
785 AAGAAAGAGAAAGAAAGAAATCTGCACTGGATCAATCATGGAGATTGGAAGAGAAAG 843
QY 879 AAAAGAACTGCCGGAACAGACTACTGGCTACAGCTCGAAATTTATTTGAAATTTATAACC 938
Db |||||
844 AAAAGAAATGGCCG-ACAGACTACTGGCTACAGCTCGAAATTTATTTGCAAAATTTATAACC 902
QY 939 AAGAACTGGGAGAGAAATATCATAGAAAGGCTATTTGTAAGGAAGTAAATTGACAAA 998
Db |||||
903 AA-AAACTGGGAAAAAATATCCAAAAAAGGGAACCGTAAAGGAGATATTTTGGAAA 961
QY 999 TATACAGCTGTTG 1011
Db |||||
962 AATATAAGCGCTG 974

RESULT 9

AL558811
LOCUS AL558811 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
DEFINITION Homo sapiens cDNA clone CS0DJ015Y19 5-PRIME, mRNA sequence.
ACCESSION AL558811
VERSION AL558811.3 GI:46184200
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 846)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On Feb 15, 2001 this sequence version replaced gi:31280609.
Contact: Genoscope
Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
5543.r

For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?8=CS0DU015AE100P1&c=5543.r.

FEATURES

source

1. .846
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DU015Y119"
/cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
/cell_line="JURKAT"
/clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT
10-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and EcoR V sites of the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 54.4%; Score 831; DB 1; Length 846;
Best Local Similarity 99.6%; Pred. No. 1.9e-179;
Matches 842; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 67 TGGGGAGTCGGATTTCTTACTCCCAAGGCTATCGCCACAGATCAAGTCCAAAGGGG 126
DB 1 TGGGGAGTCGGATTTCTTACTCCCAAGGCTATCGCCACAGATCAAGTCCAAAGGGG 60

QY 127 TGCAGAGCTACGCTGGTATTGCGAGATGTGCCAGAGCAGTGC CGGACGAGAATGGCT 186
DB 61 TGCAGAGCTACGCTGGTATTGCGAGATGTGCCAGAGCAGTGC CGGACGAGAATGGCT 120

QY 187 TTAAGTGTCTATTGATGTCGAAATCTCATCAGAGACACTATTGCTGGCTTCAGAAATC 246
DB 121 TTAAGTGTCTATTGATGTCGAAATCTCATCAGAGACACTATTGCTGGCTTCAGAAATC 180

QY 247 CTCAGCAGTTTATGGATTATTTTCAGAGGAATTCGGAATGACTTTCTAGAACTTCTCA 306
DB 181 CTCAGCAGTTTATGGATTATTTTCAGAGGAATTCGGAATGACTTTCTAGAACTTCTCA 240

QY 307 GGAGACGCTTTGGCAGCTAAAGGCTCCACAAACATTTGTCTACAACGAATACATCAGCC 366
DB 241 GGAGACGCTTTGGCAGCTAAAGGCTCCACAAACATTTGTCTACAACGAATACATCAGCC 300

QY 367 ACCGAGAGCATCCACATGAATGCCACTCAGTGGGAACTCTGACTGATTTTACTAAGT 426
DB 301 ACCGAGAGCATCCACATGAATGCCACTCAGTGGGAACTCTGACTGATTTTACTAAGT 360

QY 427 GGCTGGCAGAGAGGCTGTGCAAGTGGACGAGACACCAAGGCTGGTATATTTCAGT 486
DB 361 GGCTGGCAGAGAGGCTGTGCAAGTGGACGAGACACCAAGGCTGGTATATTTCAGT 420

QY 487 ACATAGACAGGGACCCAGAACTATCCCGCGCACTGGAACTGGAGAAAAAGAAAGC 546
DB 421 ACATAGACAGGGACCCAGAACTATCCCGCGCACTGGAACTGGAGAAAAAGAAAGC 480

QY 547 AGGACCTTGATGAAGAAAAAACTCGCCAAATTTATTGAAGACCAAGTGAAGAGGCC 606
DB 481 AGGACCTTGATGAAGAAAAAACTCGCCAAATTTATTGAAGACCAAGTGAAGAGGCC 540

QY 607 TGGAGGGAAGACAGAGGCTCCTACTTTTACGGAATTAAGCAGAGAAATCATCAAG 666
DB 541 TGGGGGGAGAGAAACAGAGGCTCCTACTTTTACGGAATTAAGCAGAGAAATCATCAAG 600

QY 667 AGAAAGTCACCTTTAATTTAGTAAAGGAGCATGTAGCTCATCCGGAGCAACATCTTCCA 726
DB 601 AGAAAGTCACCTTTAATTTAGTAAAGGAGCATGTAGCTCATCCGGAGCAACATCTTCCA 660

QY 727 AGTCAAGTACTCTGGACCGAGTGCATGAAGACGATAGGAAGTTTCAGCATCAGTGAAC 786
DB 661 AGTCAAGTACTCTGGACCGAGTGCATGAAGACGATAGGAAGTTTCAGCATCAGTGAAC 719

QY 787 GAAAGAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAAAGAAAAAAGAAATCTGCAC 846
DB 720 GAAAGAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAAAGAAAAAAGAAATCTGCAC 779

QY 847 TCGATCAATCATGAGATTGAAGAGGAGAAAGAAAGAACTCCCGAACAGACTACTGGC 906
DB 780 TCGATCAATCATGAGATTGAAGAGGAGAAAGAAAGAACTCCCGAACAGACTACTGGC 839

QY 907 TACAGCC 913
DB 840 TACAGCC 846

RESULT 10
BO221694
LOCUS
DEFINITION
AGENCOURT_7549812 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6057750
5', mRNA sequence.
BO221694
VERSION
BO221694.1 GI:20403094
EST.
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1. (bases 1 to 882)
NIH-MGC <http://mgs.nci.nih.gov/>.
NATIONAL INSTITUTES OF HEALTH, MAMMALIAN GENE COLLECTION (MGC)
UNPUBLISHED (1999)
AUTHORS
TITLE
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTF/Gazdar
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM13321 row: 1 column: 07
High quality sequence stop: 581.
Location/Qualifiers
1. .882
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6057750"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 68"
/notes="Organ: lung; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

ORIGIN

Query Match 52.4%; Score 800.4; DB 3; Length 882;
Best Local Similarity 99.6%; Pred. No. 1.9e-172;
Matches 812; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 36 GTCCAGAAAGTGTATCGCTGCCGTGGTCCCATCGGGAAGTCGGAATTTCTTACTCCCAAG 95
DB 1 GTCCAGAAAGTGTATCGCTGCCGTGGTCCCATCGGGAAGTCGGAATTTCTTACTCCCAAG 60

QY 96 GCTATGCCCAACAGGATCAAGTCCAGGGGCTCCAGAGCTACGCTGGTATTCGCCAGATG 155
DB 61 GCTATGCCCAACAGGATCAAGTCCAGGGGCTCCAGAGCTACGCTGGTATTCGCCAGATG 120

QY	156	TGCCAGAACGAGTCGCGGACGAGAAATGGCTTTAAGTGTCAATGTATGTCCGAAATCTCAT	215		
Db	121				
QY	216	CAGAGACAACTATTGCTGGCTTCAGAAAAATCCTCAGCAGTTTATGATTTATTTTCAGAG	275		
Db	181				
QY	276	GAATTCGGAATGACTTTCTAGAACTTCTCAGGAGACGCTTTGGCACTAAAAAGGGTCCAC	335		
Db	241				
QY	336	ACACACATTTGTCTACACGAATAATCAGCCACCGAGAGACATCCACATGAATGCCACT	395		
Db	301				
QY	396	CAGTGGGAAAATCTGACTGATTTTACTAAGTGGCTGGGCAGAGAAAGGCTTGTGCAAAAGTG	455		
Db	361				
QY	456	GACGAGACACAAAAGGCTGGTATATTCAAGTACATAGACAGGACCCAGAAACTATCCGC	515		
Db	421				
QY	516	CGGCAACTGGAACTGGAGAAAAGAAAAGCAGGACCTTGATGATGAAGAAAAGTGGC	575		
Db	481				
QY	576	AAATTTATTGAAGAGCAAGTGAAGAGGCTTGGAAAGGAAGCAAGAGAGTCCCTACT	635		
Db	541				
QY	636	TTTACGGAATTAAAGCAGAGAAAATGATGAAGAGAAAGTCACTTTAATTGATGAAGGA	695		
Db	601				
QY	696	GCATGTAGTCTATCCGGAGCAACATCTTCCAAGTCAAGTACTCTGGACCGAGTGCACGTG	755		
Db	661				
QY	756	AAGACGATAGGAAGTTTCAGCATCAGTGAACGAAAAGAAATCTTCCAGAGCTCAACTCAG	815		
Db	721				
QY	816	TCTAAAGAAAAG-AAGAAAAGAAATCTGCACGTGG	849		
Db	781				
RESULT 11	AY412520	1154 bp DNA linear	GSS 16-DEC-2003		
LOCUS	Mus musculus KIN gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.				
DEFINITION					
ACCESSION	AY412520				
VERSION	AY412520.1	GI:39768485			
KEYWORDS	GSS.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidae; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 1154) Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.				
TITLE	Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios				
JOURNAL	Science 302 (5652), 1960-1963 (2003)				
PUBLISHED	14671302				
REFERENCE	2 (bases 1 to 1154)				
AUTHORS	Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,				

TITLE	JOURNAL		
COMMENT	This sequence was made by sequencing genomic exons and ordering them based on alignment.		
FEATURES	Location/Qualifiers		
source	1..1154		
gene	/organism="Mus musculus" /mol_type="genomic DNA" /db_xref="taxon:10090" <1..>1154 /gene="KIN" /locus_tag="HCM4584"		
ORIGIN			
Query Match	50.2%;	Score 767.8;	DB 10; Length 1154;
Best Local Similarity	76.7%;	Pred. No. 5.9e-165;	
Matches	907;	Conservative 0;	Mismatches 247; Indels 28; Gaps 2;
QY	66	ATGGGGAAGTCGGATTTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCCAAGGGG	125
Db	1		
QY	126	CTGCAGAAAGCTACGCTGGTATTTGCCAGATGTGCCAGAAAGCAGTGC CGGAGCAGAAATGCG	185
Db	61		
QY	186	TTTAAGTGTCTATGTATGTCGGAATCTCATCAGAGACAATCTATTGCTGGCTTCAGAAAT	245
Db	115		
QY	246	CCTCAGCAGTTTATGGATTATTTTTCAGAGGAAATCCGAAATGACTTTCTAGAACTTCTC	305
Db	159		
QY	306	AGGAGACGCTTTGGCATTAAAGGGTCCACAACAATGTTCTACAAAGAATACATCAGC	365
Db	219		
QY	366	CACGAGAGCAGCATCCACATGAATGCCACTCAGTGGGAACTCTGACTGATTTTACTAAG	425
Db	279		
QY	426	TGGCTGGGCAGAGAGGCTTTGTCAAAAGTGGACGAGACACCAAAAGGCTGGTATATTAG	485
Db	339		
QY	486	TACATAGACAGGACCCAGAACTATCCGCGGCACTTGGAACTGGAGAAAAGAAAAAG	545
Db	399		
QY	546	CAGGACCTTGATGAAGAAAAGTCCCAAAATTTATTGAAGAGCAAGTGAAGAGAGC	605
Db	459		
QY	606	CTGGAAGGGAAGGAAACAGGAGGTCCCTACTTTTACGGAATTAAGCAGAGAAAATGATGA	665
Db	519		
QY	666	GAGAAAGTCAGTTTAAATTTGAGTAAAGAGCATGTAGCTCATCCGGAGCAACATCTTCC	725
Db	579		
QY	726	AAAGTCAAGTACTCTGGGACCGAGTGCCTGAAGCAGATAGGAAGTTTACAGCATCAGTGA	785
Db	639		
QY	786	CGAAAAGAAATCTTCCAGAGCTCAACTCAGTCTTAAGAAAAAGAAAAAGAAATCTGCA	845
Db	699		

Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
Adams, M.D. and Cargill, M.
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
This sequence was made by sequencing genomic exons and ordering
them based on alignment.
Location/Qualifiers
1..1154
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..>1154
/gene="KIN"
/locus_tag="HCM4584"

DEFINITION	603391378F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5406499 5', mRNA sequence.
ACCESSION	BI862461
VERSION	BI862461.1 GI:16003208
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 842)
TITLE	NIH-MGC http://mgs.nci.nih.gov/
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT	Unpublished (1999)
	Contact: Robert Strausberg, Ph.D.
	Email: cgaabs-r@mail.nih.gov
	Tissue Procurement: DCTP/DTP
	cDNA Library Preparation: Life Technologies, Inc.
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
	DNA sequencing by: Incyte Genomics, Inc.
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at:
	http://image.llnl.gov
	Plate: LLMI2036 Row: 1 Column: 20
	High quality sequence stop: 834.
FEATURES	Location/Qualifiers
source	1..842
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/clone="IMAGE:5406499"
	/tissue_type="mammary adenocarcinoma, cell line"
	/lab_host="DH10B (phage-resistant)"
	/clone_lib="NIH MGC 87"
	/notes="Organ: breast; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.383 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."
ORIGIN	
Query Match	48.7%; Score 744.6; DB 3; Length 842;
Best Local Similarity	98.8%; Pred. No. 1.2e-159;
Matches 813; Conservative	0; Mismatches 4; Indels 6; Gaps 6;
QY	27 AGAAGTGGGTCAGAAAGTATCGCTGCC-GTGGTCGCCATGGGGAAGTCGGATTTCCT 85
DB	10 AGTACTGGGTCAGAAAGTATCGCTGCCGGTGGTCGCCATGGGGAAGTCGGATTTCCT 69
QY	86 TACTCCCAAGGCTATGCCAACAGGATCAAGTCCAGGGGCTGCAGAGCTACGCTGGTA 145
DB	70 TACTCCCAAGGCTATGCCAACAGGATCAAGTCCAGGGGCTGCAGAGCTACGCTGGTA 129
QY	146 TTGCCAGATGTCAGAAAGCAGTGCAGGACAGAAATGGCTTTAAAGTGCTATTGTATGTC 205
DB	130 TTGCCAGATGTCAGAAAGCAGTGCAGGACAGAAATGGCTTTAAAGTGCTATTGTATGTC 189
QY	206 CGAATCTCATCAGAGACAACTATTGTGGCTTCAGAAAATCCTCAGCAGTTATGGATTA 265
DB	190 CGAATCTCATCAGAGACAACTATTGTGGCTTCAGAAAATCCTCAGCAGTTATGGATTA 249
QY	266 TTTTTCAGAGGAATTCGAAATGACTTTCTAGAACTTCTCAGAGAGCGTTTGGCACTAA 325
DB	250 TTTTTCAGAGGAATTCGAAATGACTTTCTAGAACTTCTCAGAGAGCGC-TTGGCACTAA 308
QY	326 AAGGGTCCACAAACAACTCTTACACAGTAATCATCAGCCACCGAGAGCACATCCACAT 385
DB	309 AAGGGTCCACAAACAACTCTTACACAGTAATCATCAGCCACCGAGAGCACATCCACAT 368
QY	386 GAATGCCACTCAGTGGGAAACTCTGACTGATTTTACTAAGTGGCTGGGCAGAGAAGGCTT 445
DB	369 GAATGCCACTCAGTGGGAAACTCTGACTGATTTTACTAAGTGGCTGGGCAGAGAAGGCTT 428
QY	446 GTGCAAGTGGAGAGACACCAAAAGGCTGGTATATTAGTATACATAGACAGGAGCCAGGA 505
DB	429 GTGCAAGTGGAGAGACACCAAAAGGCTGGTATATTAGTATACATAGACAGGAGCCAGGA 488
QY	506 AACTATCCCGCGCAACTGGAACTGGAGAAAAAGAAAGACAGGACCTTTGATGATGAAGA 565
DB	489 AACTATCCCGCGCAACTGGAACTGGAGAAAAAGAAAGACAGGACCTTTGATGATGAAGA 548
QY	566 AAAAACTGCCAAATTTATTGAAAGACAAAGTGAAGAGCCCTTGGAGGAAAGAAACAGGA 625
DB	549 AAAAACTGCCAAATTTATTGAAAGACAAAGTGAAGAGCCCTTGGAGGAAAGAAACAGGA 608
QY	626 GTGCCCTACTTTTACGGAATTAAGCAGAGAA-AAATGATGAAGAGAAAGTCAGTTTAATT 684
DB	609 GTGCCCTACTTTTACGGAATTAAGCAGAGAA-AAATGATGAAGAGAAAGTCAGTTTAATT 668
QY	685 TGAGTAAAGGAGCATGTAGCTCATCGGAGCAACATCTTCCAAAGTC-AAATGATCTCTGGGA 743
DB	669 TGAGTAAAGGAGCATGTAGCTCATCGGAGCAACATCTTCCAAAGTC-AAATGATCTCTGGGA 728
QY	744 CCGAGTGCCT-AGAGACCATAGGAAGTTTCAGATCAGTGAACGAAAAAGAAATCTTCCCA 802
DB	729 CCGAGTGCCTAGGAAGCAGATAGGAAGTTTCAGATCAGTGAACGAAAAAGAAATCTTCCCA 788
QY	803 GAGCTCAACT-CAGTCTAAAGAAAGAAAGAAAGAAATCTGC 844
DB	789 GAGCTCAACTCCAGTCTTAAAGAAAGAAAGAAAGAAATCTGC 831
RESULT 14	
LOCUS	CO725073 936 bp mRNA linear EST 27-JUL-2004
DEFINITION	ILLUMIGEN_MQ_26923 Katze_MBR Macaca mulatta cDNA clone
	IBIWI:26022 57, similar to Bases 105 to 936 highly similar to human
	KIN (Hs.123647), mRNA sequence.
ACCESSION	CO725073
VERSION	CO725073.1 GI:50705382
KEYWORDS	EST.
SOURCE	Macaca mulatta (rhesus monkey)
ORGANISM	Macaca mulatta
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Cercopitheidae; Cercopitheciniae; Macaca.
REFERENCE	1 (bases 1 to 936)
AUTHORS	Magness,C.L., Fellin,P.C., Thomas,M.J., Korth,M.J., Agy,M.B., Proil,S.C., Fitzgibbon,M., Scherer,C.A., Miner,D.G., Katze,M.G. and Iadonato,S.P.
TITLE	Analysis of the Macaca mulatta transcriptome and the sequence divergence between Macaca and human
JOURNAL	Genome Biol. 6 (7), R60 (2005)
PUBMED	15998449
COMMENT	Contact: C. Magness Illumigen Biosciences Inc. 2203 Airport Way S, Suite 450, Seattle, WA 98134, USA Tel: 2063780400 Fax: 2063780408 Email: cmagness@illumigen.com Sequenced on 2004.03.11. 732 Q20 bases. Library Preparation: Prof. Michael Katze Lab at University of Washington DNA Sequencing: Illumigen Biosciences Inc. For further information, see http://www.macaque.org PCR Primers FORWARD: CCTCTACTAAAGGAACAAAA BACKWARD: CACTATAGGCGGAATTTGGTA Insert Length: 936 Std Error: 0.00 Plate: CL000184 row: B column: 01 Seq primer: CCTCTACTAAAGGAACAAAA POLYA=yes. Location/Qualifiers
FEATURES	Location/Qualifiers
source	1..936
	/organism="Macaca mulatta"
	/mol_type="mRNA"
	/strain="Indian"

RESULT 15	CD109094	890 bp	mRNA	linear	EST 15-MAY-2000
LOCUS	CD109094				
DEFINITION	AGENCOURT_14019646 NIH_MGC_179 Homo sapiens cDNA clone				
	IMAGE:30366888 5', mRNA sequence.				
ACCESSION	CD109094	1	GI:30762268		
VERSION	CD109094.1				
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 890)				
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: sgabbs-r@mail.nih.gov Tissue Procurement: Dr. Michael Brownstein cDNA Library Preparation: Invitrogen Corp DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: NDAM427 row: d column: 01 High quality sequence stop: 653.				
FEATURES	Location/Qualifiers				
source	1. .890 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:3036688" /issue_type="Pituitary" /lab_host="DH10B-Ton A (T1 and T5 phage resistances)" /clone_lib="NIH_MGC_179" /note="Organ: Brain; Vector: PCMV-SPORT6.1; Site: 1: EcoRV (destroyed); Site: 2: NotI; Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.1 kb. Library was constructed by (Invitrogen). Note: this is a NIH_MGC Library."				
ORIGIN	Query Match 48.6%; Score 742.6; DB 6; Length 890; Best Local Similarity 88.7%; Pred. No. 3.4e-159; Matches 867; Conservative 0; Mismatches 14; Indels 96; Gaps 2				
Qy	27 AGAATCGGGTCCAGAAAGTGATCGCTGCCGTGGTCCCATCGGAGCTCGGAAGTCGGATTTCTTT 86				
Db	9 AGTACTGGGGTCCAGAAAGTGATCGCTGCCGTGGTCCCATCGGAGCTCGGAAGTCGGATTTCTTT 68				
Qy	87 ACTCCCAAGGCTATCGCCCAACAGGATCAAGTCCAAAGGGCTGCAGAAAGCTACGCTGGTAT 146				
Db	69 ACTCCCAAGGCTATCGCCCAACAGGATCAAGTCCAAAGGGCTGCAGAAAGCTACGCTGGTAT 128				
Qy	147 TGCCAGATGTGCCAGAAAGCAGTCCCGGACGAGAATGGCTTTAAGTGTCAATTGTATGTCC 206				
Db	129 TGCAGATGTGCCAGAAAGCAGTCCCGGACGAGAATGGCTTTAAGTGTCAATTGTATGTCC 161				
Qy	207 GAATCTCATCAGACACAACTATTGCTGGCTTCAGAAATCCTCAGCAGTTTATGGATTAT 266				
Db	162 -----GGAATTCGAAATGACTTTCTAGAACTTCTCAGGAGACGCTTTGGCACTAAA 326				
Qy	267 TTTTTCAGGAAATTCGAAATGACTTTCTAGAACTTCTCAGGAGACGCTTTGGCACTAAA 326				
Db	162 -----GGAATTCGAAATGACTTTCTAGAACTTCTCAGGAGACGCTTTGGCACTAAA 213				
Qy	327 AGGTCACCAACCAACATTTGCTACAACGAATACATCAGCCACCGAGACACATCCACATG 386				
Db	214 AGGTCACCAACCAACATTTGCTACAACGAATACATCAGCCACCGAGACACATCCACATG 273				

Qy	387	AATGCCACTCAGTGGGAACTCTGACTGATTTTACTAAGTGGCTGGCAGAGAGGCTTG	446
Db	274	AATGCCACTCAGTGGGAACTCTGACTGATTTTACTAAGTGGCTGGCAGAGAGGCTTG	333
Qy	447	TGCANAAGTGGACGAGACACCAAAAGGCTGGTATATTTCAGTACATAGACAGGGACCCAGAA	506
Db	334	TGCANAAGTGGACGAGACACCAAAAGGCTGGTATATTTCAGTACATAGACAGGGACCCAGAA	393
Qy	507	ACTATCCGCCGGCAACTGGAACTGGAGAAAAGAAAAAGCAGGACCTTGTATGATGAAGAA	566
Db	394	ACTATCCGCCGGCAACTGGAACTGGAGAAAAGAAAAAGCAGGACCTTGTATGATGAAGAA	453
Qy	567	AAAACCTGCCAAATTTATTGAGAGCAAGTGAAGAGGCTGGAAGGGAAGGAACAGGAG	626
Db	454	AAAACCTGCCAAATTTATTGAGAGCAAGTGAAGAGGCTGGAAGGGAAGGAACAGGAG	513
Qy	627	GTCCCTACTTTTACGGAAATTAAGCAGAGAAAATGATGAAGAGAAAGTCAAGTTTAAATTTG	686
Db	514	GTCCCTACTTTTACGGAAATTAAGCAGAGAAAATGATGAAGAGAAAGTCAAGTTTAAATTTG	573
Qy	687	AGTAAAGGAGCATGTAGCTCATCCGGAGCAACATCTTCCAAGTCAAAGTCAAGTACTCTGGGACCG	746
Db	574	AGTAAAGGAGCATGTAGCTCATCCGGAGCAACATCTTCCAAGTCAAAGTCAAGTACTCTGGGACCG	633
Qy	747	AGTGCACCTGAGACGATAGGAAGTTCAGCATCAGTGAACGAAAAGAAATCTTCCCAGAGC	806
Db	634	AGTGCACCTGAGACGATAGGAAGTTCAGCATCAGTGAACGAAAAGAAATCTTCCCAGAGC	693
Qy	807	TCAACTCAGTCTAAAGAAAAGAAAAGAAAATCTGCCTGGATGAAATCATGGAGATT	866
Db	694	TCAACTCAGTCTAAAGAAAAGAAAAGAAAATCTGCCTGGATGAAATCATGGAGATT	753
Qy	867	GAAAGAGAAAAGAAAAGAACTGCCCGAACAGACTACTGGCTACAGCC-TGAAAATTATTGT	925
Db	754	GAAAGAGAAAAGAAAAGAACTGCCCGAACAGACTACTGGCTACAGCC-TGAAAATTATTGT	813
Qy	926	GAAAAATTATACCAAGAACTGGGAGAGAAATATCATAGAAAAAGGCTATTGTTAAGGA	985
Db	814	GAAAAATTATACCAAGAACTGGGAGAGAAATATCCTTAAGAAAAAGGCTATTGTTAAGGA	873
Qy	986	AGTAATTGACAAATATA 1002	
Db	874	AGTATTGACAAATATA 890	

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:00:02 ; Search time 316.435 Seconds
(without alignment)
8583.479 Million cell updates/sec

Title: US-09-555-529-1

Perfect score: 1528
Sequence: 1 ctgaattcagggcgctg.....atgtattataaaaaaaaaa 1528

Scoring table: IDENTITY_NUC
Gapop 10.0', Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/1 COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5 COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/H COMB.seq.*
6: /cgn2_6/ptodata/1/ina/PP COMB.seq.*
7: /cgn2_6/ptodata/1/ina/RE COMB.seq.*
8: /cgn2_6/ptodata/1/ina/RE COMB.seq.*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	151.4	9.9	23645	US-09-949-016-13916	Sequence 13916, A
C 2	123.6	8.1	538	US-09-270-767-379	Sequence 379, App
C 3	123.6	8.1	538	US-09-270-767-15661	Sequence 15661, A
C 4	81	5.3	825	US-09-248-796A-5938	Sequence 5938, Ap
C 5	79.6	5.2	7218	US-08-232-463-14	Sequence 14, Appl
C 6	53.2	3.5	25951	US-09-949-016-16194	Sequence 16194, A
C 7	51	3.3	1141	US-09-806-708B-22	Sequence 22, Appl
C 8	50.6	3.3	1141	US-09-806-708B-22	Sequence 22, Appl
C 9	46.2	3.0	147382	US-09-949-016-14624	Sequence 14624, A
C 10	45.8	3.0	209210	US-09-949-016-15094	Sequence 15094, A
C 11	45.4	3.0	612	US-09-902-540-1357	Sequence 1357, Ap
C 12	45.4	3.0	5181	US-08-257-073-10	Sequence 10, Appl
C 13	44.6	2.9	1850	US-08-617-860B-32	Sequence 32, Appl
C 14	44.6	2.9	4098	US-08-605-106-4	Sequence 4, Appl
C 15	44.6	2.9	137394	US-09-949-016-13872	Sequence 13872, A
C 16	44.6	2.9	137743	US-09-949-016-12178	Sequence 12178, A
C 17	44.6	2.9	640681	US-09-790-988-1	Sequence 1, Appl
C 18	44.4	2.9	37335	US-09-949-016-17132	Sequence 17132, A
C 19	44	2.9	4920	US-09-269-874A-1	Sequence 1, Appl
C 20	43.6	2.9	114139	US-09-949-016-16536	Sequence 16536, A
C 21	43	2.8	612	US-09-902-540-1357	Sequence 1357, Ap
C 22	42.6	2.8	227750	US-09-949-016-17175	Sequence 17175, A
C 23	42.2	2.8	168174	US-10-071-411A-63	Sequence 63, Appl
C 24	42.2	2.8	168273	US-10-071-411A-2	Sequence 2, Appl

C 25	41.8	2.7	6243	2	US-09-056-075-1	Sequence 1, Appl
C 26	41.6	2.7	2181	3	US-09-662-254B-40	Sequence 40, Appl
C 27	41.6	2.7	2518	3	US-09-433-699-3	Sequence 3, Appl
C 28	41.6	2.7	2527	3	US-09-949-016-4169	Sequence 4169, Ap
C 29	41.6	2.7	2637	3	US-10-104-047-681	Sequence 681, App
C 30	41.6	2.7	50000	3	US-09-662-254B-24	Sequence 24, Appl
C 31	41.6	2.7	91062	3	US-09-949-016-13019	Sequence 13019, A
C 32	41.4	2.7	91062	3	US-09-734-674-3	Sequence 3, Appl
C 33	41.4	2.7	202001	3	US-10-274-990-3	Sequence 89871, A
C 34	41.2	2.7	601	3	US-09-949-016-89871	Sequence 93431, A
C 35	41.2	2.7	601	3	US-09-949-016-93431	Sequence 162984, A
C 36	41.2	2.7	601	3	US-09-949-016-162984	Sequence 14429, A
C 37	41.2	2.7	96987	3	US-09-949-016-14429	Sequence 43, Appl
C 38	41	2.7	2304	3	US-09-662-254B-43	Sequence 24, Appl
C 39	41	2.7	50000	2	US-09-662-254B-24	Sequence 3, Appl
C 40	40.8	2.7	3867	2	US-07-876-280-3	Sequence 3, Appl
C 41	40.8	2.7	3867	2	US-07-675-772-3	Sequence 3, Appl
C 42	40.8	2.7	3867	2	US-08-063-170-3	Sequence 3, Appl
C 43	40.8	2.7	3867	2	US-08-158-232-3	Sequence 3, Appl
C 44	40.8	2.7	3867	2	US-08-304-626-3	Sequence 3, Appl
C 45	40.8	2.7	3867	2	US-08-316-301A-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-09-949-016-13916/c
; Sequence 13916, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13916
; LENGTH: 23645
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13916

Query Match	9.9%	Score 151.4;	DB 3;	Length 23645;
Best Local Similarity	99.3%	Pred. No. 7.2e-28;		
Matches 152;	Conservative	0;	Mismatches 1;	Indels 0; Gaps 0;
QY	27	AGAACTGGGGTCCAGAAAGTATCGCTCGGTGGTCCGATCGGGAAGTCGGATTTCTT	86	
DB	1802	AGTACTGGGGTCCAGAAAGTATCGCTCGGTGGTCCGATCGGGAAGTCGGATTTCTT	1743	
QY	87	ACTCCCAAGGCTATCGCCAAAGGATCAAGTCCAGGGGCTCCAGAGCTACGCTGGTAT	146	
DB	1742	ACTCCCAAGGCTATCGCCAAAGGATCAAGTCCAGGGGCTCCAGAGCTACGCTGGTAT	1683	
QY	147	TGCCAGATGTCAGAGCAGTCGCCGGGACGAG	179	
DB	1682	TGCCAGATGTCAGAGCAGTCGCCGGGACGAG	1650	

RESULT 2
US-09-270-767-379/c
; Sequence 379, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:


```
||||| |||||
Db 20110 ATATATTATAATA 20097

RESULT 7
US-09-806-708B-22/c
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22
Query Match 3.3%; Score 51; DB 3; Length 1141;
Best Local Similarity 10.5%; Pred. No. 0.006;
Matches 110; Conservative 382; Mismatches 555; Indels 4; Gaps 2;

Qy 329 GGTCCACACACATTTGTCTCAAGAAATACATCAGCCAGCAGACACATCCACATGAA 388
Db 1109 GKDWRMDATKWSATGTATWTHAKRGATCMCYWYWTGTNRRCWRTYAMRTWYTRSNAN 1050
Qy 389 TGCACACTCAGTGGGAACTCTGACTGATTTTACTAAGTGGCTGGCAGAGAGGCTTGTG 448
Db 1049 WSCATGKBMWMTKMYATKYRTAWYAMCAWRNNNMWCATNGYAKSCATNNNAWYATTRWA 990
Qy 449 CAAAGTGACGACGACACACAAAGGCTGTATATTTCAGTACATAG---ACAGGGACCCAGA 505
Db 989 AYAAKAWARWAGNNRMRYGAAAGNKWGCMAAMATMBGWATAGWCMNNNNNNNTDVR 930
Qy 506 AACTATCCGCGGCAACTGGAACTGGAGAAAGAAAGACGACCTTGATGATGAAGA 565
Db 929 MAMKAKNNNNNAYWTCYNRAATNNKQATWMMKWTGHAHSKRTRHTRTCRTKYNN 870
Qy 566 AAAAAGTCCCAATTTATTGAAGACCAAGTGAAGAGGCTGGAGGGAAGGACAGGA 625
Db 869 NNNARTVYVYHHAARRMNNATWTRTNNNNNNNNNACRNTRTWABWKHSWCMNNNNNN 810
Qy 626 GGTCCCTACTTTTACGGAATTAAGCAGAGAAATGATGAAGAGAAAGTCAAGTTTAATT 685
Db 809 NNNNNNTWCHYTTANABCBYRANNNNAARWARTCNMYHAAVTTTHDWCYKWTWNTY 750
Qy 686 GAGTAAAGGACATGTAGCTCATCCGGAGCAACATCTTCCAAAGTCAAGTACTCTGGGACC 745
Db 749 WDMTMTTBTTRNWT-TSTNMTNNNNNMWACTNNNNNNNNNNNNNNNNNNNNNNNNNNNN 691
Qy 746 GAGTGCATCAGACGATAGGAGTTCAGCATCAGTGAACGAAAGAAATCTTCCAGAG 805
Db 690 DARRNTNTVNRMMWMTKTRWSTTRRHHTYGTATNNNNNNNNNNNNNNNNNNNNNNNNNN 631
Qy 806 CTCAACTCAGTCTAAAGAAAGAAAGAAAGAAATCTGCACTGGATGAAATCATGGAGAT 865
Db 630 MTMRWTKGDGWTVRKKYKRDCTTYDVWADSWWYHAYWRCRDVYTRNTYCKSY 571
Qy 866 TGAAGAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 925
Db 570 AHSYWYWSNNNAWYRRYSARNWSSMARWTTTRNNNNNNNNNNNNNNNNNNNNNNNNNN 511
Qy 926 GAAATATTATACCAAGAAACTGGGAGAGAAATATATCAAGAAAGAAAGGCTATTGTTAAG 985
```

```

Db 510 RYYWKKRWABBTVTYYDSMCNAKSMWRGNNGNMMRAMKMMWAAANNNDAGAMDHWTYMGNNTM 451
Qy 986 AGTAATTTGACAAATATATACAGCTGTGTGAAGATGATTGATTTCTGGAGACAAGCTGAAACT 1045
Db 450 WMRRAWKMNMWACRRAYCCNNNNNRACVHHKHWRWRTWKYMWKAACNNNNNBKAMTMRV 391
Qy 1046 TGACGAGACTCATTTAGAGACAGTAATTCAGCACCCAGGAAAAAGAAATTTCTAGTTTTAAA 1105
Db 390 AMMYSDTTNTDMMWTSDBWHWYTVDTYMMRAMNNNNNNNNNNNNNNNNNNNNNNNNNNNN 331
Qy 1106 TGGAGGCTACAGAGGAATGAAGTCCCTAGATCCATCATCAATCAGAGAGACTTTTTCAGC 1165
Db 330 THCTYGNNTWGSAYBMAASMSMAAGASBNVTYNCWMTYMGKTMNTNNNNNNKAWYRTK 271
Qy 1166 TACTATCGTCATTGAAACTGGCCCTTTTAAAGAGCAGCAGAGTGTGAAGAAATTTCAATATGA 1225
Db 270 TVAWCNRRYYDITAVWBTBKENYKYCYAYBMYBYMGKHHBWMWRRAHBHSNNNNWVKCR 211
Qy 1226 AGACATTTCTAACTTCGCTGAGTTTGAAAAATTTGTTTAAACAATACATATAAATCTTAAAG 1285
Db 210 NKYMVSWHYHAMRYBKWABAVGCNNNNNNKDRMAHHHCATNNNNNNNNNNNNNNNNNNNN 151
Qy 1286 CATCAAAATTCGTGTTCCCAAGGCATTATGAGACTCTACTGCTGTGTTAGGCTATATCTTT 1345
Db 150 AWTNNKTABRDDHBAHVKTYYWRYDYWCAMCWNNAKAKVKTAMKHMWYTDTRYVSANNT 91
Qy 1346 GTATAAAACAAACAGGTTTTTTGAAAAATATTA 1376
Db 90 GVRMMWRMCMWYWSMNNRMYRMRKYTWA 60

RESULT 8
US-09-806-708B-22
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22
```

```
Query Match 3.3%; Score 50.6; DB 3; Length 1141;
Best Local Similarity 10.6%; Pred. No. 0.0076;
Matches 62; Conservative 243; Mismatches 279; Indels 2; Gaps 1;

Qy 943 AACTGGGAGAGAAATATCATGAAGAAAGGCTATTGTTAAGGAAGTAATTGACAAATATA 1002
Db 91 ANNTSBRHYARRWKKDKMTAYBMTNKNWGTGRWRYWRWRAEDTVDDHHYVTAMNNWAT 150
Qy 1003 CAGCTGTGTGAAGATGATTCTTGAGACAAGCTGAAACTTTGACCAGACTCATTTAG 1062
Db 151 TCMCMKDDKRTWKKNNNNATGWDHDDTKYHMMNNNGCBETVIMVRYKTRDWSBKRBN 210
Qy 1063 AGACAGTAATTCAGCACCCAGGAAAAAGAAATTTCTAGTTTTTAAATGGAGGCTACAGAGAA 1122
Db 211 YGBMWKWSYDVTYYWVWDDMKCRKVRWRVTRGRMRNMYV--AMBTAAHRRYNNNGWT 268
Qy 1123 ATGAGAGTACCTAGATCCATCAATCAGAGAGACTTTTTCAGCTACTACTTCGTCATTGAAA 1182
```

Db 269 BAAATRRWNNNNNNNAKACRAKYGWGNBVBVNSTCTTWKSKTKTKVTSWANNCRAG 328
Qy 1183 CTGGCCCTTTAAAGGACGACAGGTTGAAGAAATCAATATAGAGACATTTCTAAACATTG 1242
Db 329 DANKDHKKWKAAGVYNNNNNNNNNTYKARHBBARMDVWHSAAKWHANAHAHYSRKK 388
Qy 1243 CTGAGTTTGAAGAAATTTGTAACAATACATTAATTAATCTTAAGCATCAAAATGGTGTTCG 1302
Db 389 WTBYRKTMVNNNGTTTMMKRWAMYKMDMBWGTNNNNNGRTYYGWTNKKKQWYY 448
Qy 1303 CCAAGGCATTATGAGACTCTACTGTGTAGGGTATATCTTTTGTATATAAAACAAACAGGT 1362
Db 449 KWKANCKWRNDPKTCTHNTTWWKMTYNNWYKSWTNGKSHRBAAVYTWYMWWR 508
Qy 1363 TTTTGAATAATATCTGTATAGTTGTTACGCTAAACTTTTGAGAAGAAATTAATATGCT 1422
Db 509 RYAHANNNDYWKACTYKVBVCCKWNNYAAWYTKSSMNYTSRYVRWKTNNRWRS 568
Qy 1423 CATGAGTATCAAACTATGTAATTTTGTCTTGTATTATTCTTCTTCTTGTAAATTTACT 1482
Db 569 DTRSMGRANNYARABHYGYKWNTRWBSHTWBHBRAGAHHYMBWMMYBAKCHCKWATK 628
Qy 1483 TGATGAGTTTATATCTTCTTCAATGAAGATGTTATTATAAAAAA 1528
Db 629 AKKIAGAGSSNNNNNNNNNNNNNNATCARDYYAASRWYMANA 674

RESULT 9
US-09-949-016-14624
; Sequence 14624, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 14624
; LENGTH: 147382
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(147382)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14624

Query Match 3.0%; Score 46.2; DB 3; Length 147382;
Best Local Similarity 52.3%; Pred. No. 0.72;
Matches 102; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
Qy 1333 GGTATATCTTTGTATAAACAACAGGTTTGTGAATAATATCTGTATAGTTGTCAG 1392
Db 136880 GTTATATATTTATATATAGTTATATATATATATATATATATATATATATATAT 136939
Qy 1393 CTAACTTCGAGAGATTTAATATGCTCATGAGTATCAAACTATGAAATTTGTCC 1452
Db 136940 ATATATGATATATATTAACATATTTATATATATATATATATATATATATATATAT 136999
Qy 1453 TTGTTATTTTGTTCCTTTGTAATTTTACGTATGAGTTTATATCTTCATTAAGAATGT 1512
Db 137000 TAATATATTTATGT 137059

Qy 1513 TATTATAAAAAA 1527
Db 137060 ATATATATATATA 137074
RESULT 10
US-09-949-016-15094
; Sequence 15094, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 15094
; LENGTH: 209210
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(209210)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15094
Query Match 3.0%; Score 45.8; DB 3; Length 209210;
Best Local Similarity 50.6%; Pred. No. 1;
Matches 136; Conservative 0; Mismatches 132; Indels 1; Gaps 1;
Qy 1251 TGAAAATTTGTTAAACAATACATTAATAATCTTAAGCATCAAAATGGTGTGCCAAGCA 1310
Db 112647 TGAAGTATCTTTACATGTCAGTTAATCTGATGTTTATATATCTGTCAGTATTTC 112706
Qy 1311 TTATGAGACTCTACTGTGTGTAGGGTATATCTTTTGTATAAAACAAACAGGTTTGTAAA 1370
Db 112707 TTATTAATCTTCTGTATAGTTTGTCCATTTTGAAGTGAGATATTGAAGTCTCAAAA 112766
Qy 1371 ATATTACTGTATAGTTGTTGACGTAACCTTTGAGAAGAAATTTAATATGTCCTCATGAGT 1430
Db 112767 CTAATATTTTGGAGTTGCTATTCTCTCTTTCAGTTTTCAGTTTATGTTTTCATATAT 112826
Qy 1431 ATCAAACTATGTAATTTGTCTCTCTCT-TATTTTGTCTTCTTCTTGTAAATTTACTTGATGAG 1489
Db 112827 TTGGCACTCCGTTGTTAGTCCATGATGTTTATTAATTTGTTATATCTTCTTGTGATG 112886
Qy 1490 TTTATATCTTCAATAAAGAAATGTTATTAT 1518
Db 112887 ATCCTTTTATCATTAGAAAATGCTTTGT 112915

RESULT 11
US-09-902-540-1357
; Sequence 1357, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883

```
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1357
; LENGTH: 612
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(612)
; OTHER INFORMATION: unsure at all n locations
US-09-540-1357

Query Match
Best Local Similarity 3.0%; Score 45.4; DB 3; Length 612;
Matches 205; Conservative 0; Mismatches 268; Indels 0; Gaps 0;

QY 532 AGAAAAAGAAAAAGCAGGACCTGTGATGTAAGAAAAAAGTCCCAAAATTTATTGAAGAGC 591
DB 80 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 139
QY 592 AAGTGAGAGAGCGCTGGAAGGAGGAAAGACAGGAGGTCCCTACTTTTACGGAAATTAAGCA 651
DB 140 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 199
QY 652 GAGAAATCATGACGAGAAAGTCACGTTTAAATTTGAGTAAGGAGCATGTAGCTCATCG 711
DB 200 AATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 259
QY 712 GAGCAACATCTTCCAGTCAAGTACTCTGGGACCGAGTGCACTGAAGACGATAGGAAGTT 771
DB 260 TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 319
QY 772 CAGCATCAGTGAACGAAAGAAATCTCCAGAGCTCAACTCAGTCTAAAGAAAAAGAGA 831
DB 320 AAAAAAAAAAAAAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 379
QY 832 AAAAGAAATCTGCAGTGGATGAATCATGGAGATTGAAGAGGAAAAAGAAAGAACTGCC 891
DB 380 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 439
QY 892 GAACAGACTACTGGCTCAGACCTGAAATTTATGTGAAATTTATAACCAAGAACTGGAG 951
DB 440 TTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 499
QY 952 AGAAATATCATAGAAAAAGCGTATTGTTAAGGAAGTAATTGACAAATATACA 1004
DB 500 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 552

RESULT 12
US-08-257-073-10/c
; Sequence 10, Application US/08257073
; Patent No. 5766597
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: de Taisne, Charles
; APPLICANT: Tine, John A.
; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue, 25th Floor
; CITY: New York
; STATE: New York
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,073
```

```
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION NUMBER: US 08/075,783
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/852,305
; FILING DATE: 18-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,183
; FILING DATE: 20-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Prommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2570
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-0712
; TELEFAX: (212) 840-0712
; TELEX: 425066 CURTMS
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5181 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-257-073-10

Query Match
Best Local Similarity 3.0%; Score 45.4; DB 2; Length 5181;
Matches 158; Conservative 0; Mismatches 146; Indels 5; Gaps 2;

QY 1223 TGAAGACATTTCTAAATTCGCTGAGTTTGAAATTTGTTAAACATACATATAAATCTTA 1282
DB 3483 TAAAGTTTAAATGAGATGATTCACCATTAATAATATTAAACAGTCTTTATAATGTTT 3424
QY 1283 AAGCATCAAAATTTGTTTCGCCAAGGCATTATGAGACTCTACTGTGTAGGTATATTCT 1342
DB 3423 CAATAATATTTTGTGTTTCTAATGTTTTCAGTTTCGTATTTTCAGCTCTCTTTT 3364
QY 1343 TTTGT---ATAAACAAACAGGTTTTTGAATAATTAATCT--GTATAGTTGTTTCAGCTAAA 1397
DB 3363 TTTCTTAAAGAAACACAGAAAGTTTGTATAATACATTAATGTTGGTTATTAAAGTGAATCAA 3304
QY 1398 CTTTGAGAGAAATTAATTAATGCTCATGAGGTATCAAACTATGTAATTTTGCTGTTT 1457
DB 3303 TTTTGATTCTAAATTTGTTCTTTTAAATAAGTAAGTTTTTAAATTTGCAATTTGCTTGCC 3244
QY 1458 ATTTTGTCTTCTTTGTAATTTACTTCATGAGTTTATATCTTCATTAAAGAAATGTTATTA 1517
DB 3243 AAGTTCTTTTCTTATTAATAATCTATCTAATTTTAAATTTATATATATAATAATC 3184
QY 1518 TAAAAAAA 1526
DB 3183 AATGATAA 3175

RESULT 13
US-08-617-860B-32
; Sequence 32, Application US/08617860B
; Patent No. 6133506
; GENERAL INFORMATION:
; APPLICANT: Tvpfer, R., Bautor, J., Bothmann, H., Flisak, E.,
; APPLICANT: Hvrlicke-Grandpierre, C., Klein, B., Martini, N.,
; APPLICANT: Miller, A., Schulte, W., Voetz, M., Walek, J.,
; APPLICANT: Schell, J.
; TITLE OF INVENTION: Promoters
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steinberg, Raskin & Davidson, P.C.
; STREET: 1140 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
```


Query Match	2.9%	Score 44.6	DB 3	Length 1850
-------------	------	------------	------	-------------

```
; FEATURE:
; NAME/KEY: intron IV
; LOCATION: 3012..3131
; FEATURE:
; NAME/KEY: exon V
; LOCATION: 3132..3303
; FEATURE:
; NAME/KEY: intron V
; LOCATION: 3304..3390
; FEATURE:
; NAME/KEY: exon VI
; LOCATION: 3391..3459
; FEATURE:
; NAME/KEY: intron VI
; LOCATION: 3460..3671
; FEATURE:
; NAME/KEY: exon VII
; LOCATION: 3672..3941
; FEATURE:
; NAME/KEY: Stopcodon
; LOCATION: 3942..3944
;
US-08-605-106-4

Query Match
Best Local Similarity 2.9%; Score 44.6; DB 2; Length 4098;
Matches 128; Conservative 0; Mismatches 119; Indels 2; Gaps 1;

QY 1271 ATTAAATCTTAAAGCATCAAAATCGTGTTCCCAAGGCATTATGAGACTCTACTGTGTT 1330
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
141 ATTTAAATTTTTCGGGATATATTTGTAATATTTTATGAATTTATGAAATATTTTTCGAA 200
QY 1331 AGGGTATATCTTTTGTATATAAACAACAGGTTTTTGAATAATTTACTGTATAGTTGTC 1390
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
201 ATTTTAAATATTTTAAATTTTAAATTTTAAATATATTTTAAATTTCTTTTAAAAAATAATTTT 260
QY 1391 AGCTAAACTTTGAGAGAAATTTAAATTTATGTCATGAGTATCAAACTATGTAATTTTGT 1450
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
261 AAATATTATAAATTTAGTTTTTAAATTTTTTAA--ATATTTAAATTTAGTTTTTTTAA 318
QY 1451 CTTGTTATTTTGTTCCTTTGTAAATTTACTGTAGTATATCTTCATTAAGAAT 1510
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
319 TTTTAAATATTTGTTGAATTTTAAATATTTTGGTTTTTAAATATATATTTTAAAGT 378
QY 1511 GTTATTATA 1519
Db |||||
379 TTTTAAATA 387

RESULT 15
US-09-949-016-13872/c
; Sequence 13872, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13872
; LENGTH: 137394
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
```

```
; LOCATION: (1)....(137394)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13872

Query Match
Best Local Similarity 2.9%; Score 44.6; DB 3; Length 137394;
Matches 95; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 820 AAGAAAAAGAGAAAAAGAAATCTGCACTGATGAAATCATGAGATTGAGAGAGAAAAAGA 879
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
15531 AAGAAAAAGAGAAAAAGAAAGAGAAAGAGAAAGAGAAAGAGAGAGAGAGAGAGAGAA 15472
QY 880 AAAGAACTGCCCGAACAGACTACTGGCTACAGCTGAAATTTTGTGAAAATTTATAACCA 939
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
15471 AAGGAAAGAGAAAGGAAAGAGAAAGAGAAAGAGAAAGAGAGAGAGAGAGAGAGAGAA 15412
QY 940 AGAAACTGGGAGAGAAAAATATCATAGAAAAAGGCTATTGTTAAGGAAAGTAAATTGACAAA 998
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
15411 AGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 15353

Search completed: November 27, 2005, 01:10:36
Job time : 319.435 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:00:41 ; Search time 1761.53 Seconds
(without alignment)
7173.088 Million cell updates/sec

Title: US-09-555-529-1
Perfect score: 1528
Sequence: 1 ctgaattcagcgccgctg.....atgtattataaaaaaaaaa 1528

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA.Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	490.8	32.1	591	5	US-10-106-698-1187 Sequence 1187, Ap
2	436.4	28.6	461	7	US-10-242-535A-29177 Sequence 29177, A
3	436.4	28.6	461	7	US-10-085-783A-29177 Sequence 29177, A
4	298.6	19.5	1241	10	US-11-097-143-41039 Sequence 41039, A
5	284	18.6	300	9	US-10-779-543-7646 Sequence 7646, Ap
6	247.2	16.2	3313	10	US-11-097-143-41038 Sequence 41038, A
7	209.4	13.7	4394	10	US-11-097-143-41017 Sequence 41017, A
8	209.4	13.7	4582	10	US-11-097-143-24532 Sequence 24532, A
9	205.6	13.5	777	8	US-10-424-599-25484 Sequence 25484, A
10	204.8	13.4	1713	8	US-10-739-930-2955 Sequence 2955, Ap
11	203.2	13.3	2046	8	US-10-425-115-157125 Sequence 157125, Ap
12	201.6	13.2	1614	7	US-10-425-114-2975 Sequence 2975, Ap
13	201.6	13.2	1713	7	US-10-425-114-22614 Sequence 22614, A
14	192	12.6	538	7	US-10-437-963-89433 Sequence 89433, A
15	156.8	10.3	584	7	US-10-767-701-25587 Sequence 25587, A
16	126.4	8.3	549	3	US-09-991-936-1424 Sequence 1424, Ap
17	126.4	8.3	549	9	US-10-978-245-1424 Sequence 1424, Ap
18	116	7.6	116	7	US-10-242-535A-7473 Sequence 7473, Ap
19	116	7.6	116	7	US-10-085-783A-7473 Sequence 7473, Ap
20	106.8	7.0	739	7	US-10-424-599-53324 Sequence 53324, A
21	102.4	6.7	516	4	US-09-925-065A-549989 Sequence 549989, Ap
22	100.2	6.6	270	3	US-09-294-093B-3089 Sequence 3089, Ap
23	78.4	5.1	588	4	US-09-925-065A-20834 Sequence 20834, A

24	72.6	4.8	1608	4	US-09-925-065A-707309 Sequence 707309, A
25	71.2	4.7	616	7	US-10-424-599-53326 Sequence 53326, A
26	60	3.9	60	3	US-09-908-975-6340 Sequence 6340, Ap
27	59.8	3.9	521	8	US-10-425-115-116286 Sequence 116286, A
28	59.4	3.9	12578	7	US-10-221-714A-382 Sequence 382, App
29	58	3.8	5935	5	US-10-239-676-134 Sequence 134, App
30	56.8	3.7	2985	10	US-11-097-143-41056 Sequence 41056, A
31	56	3.7	3673778	6	US-10-312-841-1 Sequence 1, Appl1
32	55.2	3.6	358	7	US-10-424-599-106049 Sequence 106049, A
33	55.2	3.6	11422	6	US-10-311-455-192 Sequence 192, App
34	55.2	3.6	11422	7	US-10-257-166-18 Sequence 18, Appl1
35	54.4	3.6	65	3	US-09-908-975-24642 Sequence 24642, A
36	54	3.5	15951	6	US-10-311-455-1653 Sequence 1653, Ap
37	54	3.5	15951	6	US-10-240-485-133 Sequence 133, App
38	53.8	3.5	1204	7	US-10-437-963-77858 Sequence 77858, A
39	53.8	3.5	6831	6	US-10-311-455-1459 Sequence 1459, Ap
40	53.4	3.5	6013	7	US-10-221-613-225 Sequence 225, App
41	53	3.5	8056	8	US-10-473-146-386 Sequence 386, App
42	52.6	3.4	113515	6	US-10-311-455-2148 Sequence 2148, Ap
43	51.8	3.4	3720	5	US-10-172-086-74 Sequence 74, Appl
44	51.8	3.4	3720	7	US-10-311-507-38 Sequence 38, Appl
45	51.8	3.4	3720	8	US-10-480-846-74 Sequence 74, Appl

ALIGNMENTS

RESULT 1

US-10-106-698-1187
; Sequence 1187, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA005PI
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 1187
; LENGTH: 591
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (587)..(587)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-1187

Query Match 32.1%; Score 490.8; DB 5; Length 591;

Best Local Similarity 97.0%; Pred. No. 1.6e-109;
Matches 512; Conservative 0; Mismatches 12; Indels 4; Gaps 1;

QY	1005	GCTGTTGTGAAGATGATTTCTCGAGACAAGCTGAAACTTGACACAGACTCATTTAGAG	1064
DB	28	GCTGCAGGAATTCGGCAGCAGTCTCGAGACAAGCTGAAACTTGACACAGACTCATTTAGAG	87
QY	1065	ACAGTAATTCAGCACCAGCAAAAGAAATTCAGTTTAAATGGAGGCTACAGAGAAAT	1124
DB	88	ACAGTAATTCAGCACCAGCAAAAGAAATTCAGTTTAAATGGAGGCTACAGAGAAAT	147
QY	1125	GAAGGTACCTAGATCCATCAATCAATGAGAGACTTTTTCAGTACTATCGTCTATTGAACT	1184
DB	148	GAAGGTACCTAGATCCATCAATCAATGAGAGACTTTTTCAGTACTATCGTCTATTGAACT	207
QY	1185	GGCCCTTTTAAAGGAGCGCAGAGTTGAAGGAAATTCATATGAAAGACATTTCTAACTTGGC	1244

```
Db 208 GSCCCTTTAAAGGACGAGAGTTGAAGAAATTCATATGAGACATTTCTAAACTTGCC 267
QY 1245 TGAGTTTGAAATTTGTTAAACAATACATTAATAATCTTAAAGCATCAAAATGGTGTGCC 1304
Db 268 TGAGTTTGAAATTTGTTAAACAATACATTAATAATCTTAAAGCATCAAAATGGTGTGCC 327
QY 1305 AAGCATTATGAGACTCTACTGTGTAGGATATATCTTTGTATATAAACAACAGGTTT 1364
Db 328 AAGCATTATGAGACTCTACTGTGTAGGATATATCTTTGTATATAAACAACAGGTTT 387
QY 1365 TTGAAATATTAATCTGTA----TAGTGTGTTACGTAAACCTTTGAGAGAAATTTAATATGT 1420
Db 368 TTGAAATATTAATCTGTAATAGTTAGTTGTTACGTAAACCTTTGAGAGAAATTTAATATGT 447
QY 1421 CTCATGAGTATCAAACTATGTAATTTGTCCTGTTATTTTGTGTTCTTTGTAATTTA 1480
Db 448 CTCATGAGTATCAAACTATGTAATTTGTCCTGTTATTTTGTGTTCTTTGTAATTTA 507
QY 1481 CTTGATGAGTTTATATCTTCATTAAGAATGTTATTATATAAAAAAAA 1528
Db 508 CTTGATGAGTTTATATCTTCATTAAGAATGTTATTATATAAAAAAAA 555

RESULT 2
US-10-242-535A-29177
; Sequence 29177, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US 10/242,535A
; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29177
; LENGTH: 461
; TYPE: DNA
; ORGANISM: Human
US-10-242-535A-29177

Query Match 28.6%; Score 436.4; DB 7; Length 461;
Best Local Similarity 99.4%; Pred. No. 2.9e-96;
Matches 459; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 59 GGTGCCCATGGGAAGTCGGATTTCTTACTCCCAAGCTATCGCCACAGGATCAAGTC 118
Db 1 GGTGCCCATGGGAAGTCGGATTTCTTACTCCCAA-GCTATCGCCAAACAGGATCAAGTC 59
QY 119 CAAGGGCTGCAGAAAGCTACGCTGTTATTCAGAGATGTGCAGAAAGCAGTCCCGGACGA 178
Db 60 CAAGGGCTGCAGAAAGCTACGCTGTTATTCAGAGATGTGCAGAAAGCAGTCCCGGACGA 119
QY 179 GAATGGCTTTAAGTGTCAATGTATGTCGGAATCTCATCAGAGACAACATATTGCTGCTTC 238
Db 120 GAATGGCTTTAAGTGTCAATGTATGTCGGAATCTCATCAGAGACAACATATTGCTGCTTC 179
QY 239 AGAAATCCTCAGCAGTTTATGGAATATTTTTCAGAGAAATTCGAAATGACTTTCTAGA 298
Db 180 AGAAATCCTCAGCAGTTTATGGAATATTTTTCAGAGAAATTCGAAATGACTTTCTAGA 239
QY 299 ACTTCTCAGGAGACGCTTTGCGACTAAAGGGTCCACAACAACATTTCTTACACGAATA 358
Db 240 ACTTCTCAGGAGACGCTTTGCGACTAAAGGGTCCACAACAACATTTCTTACACGAATA 299
QY 359 CATCAGCCACCGAGAGACATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTT 418
Db 300 CATCAGCCACCGAGAGACATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTT 359
QY 419 TACTAAGTGTGCTGGC-AGAGAAAGCTTTGCAAAAGTGGACGAGACACCAAAAGGCTGGT 477
Db 360 TACTAAGTGTGCTGGC-AGAGAAAGCTTTGCAAAAGTGGACGAGACACCAAAAGGCTGGT 419
QY 478 ATATTCTAGTACATAGACAGGACCCAGAAACTATCCGCCGCC 519
Db ATATTCTAGTACATAGACAGGACCCAGAAACTATCCGCCGCC 461
```

```
QY 359 CATCAGCCACCGAGAGACATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTT 418
Db 300 CATCAGCCACCGAGAGACATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTT 359
QY 419 TACTAAGTGTGCTGGC-AGAGAAAGCTTTGCAAAAGTGGACGAGACACCAAAAGGCTGGT 477
Db 360 TACTAAGTGTGCTGGC-AGAGAAAGCTTTGCAAAAGTGGACGAGACACCAAAAGGCTGGT 419
QY 478 ATATTCTAGTACATAGACAGGACCCAGAAACTATCCGCCGCC 519
Db 420 ATATTCTAGTACATAGACAGGACCCAGAAACTATCCGCCGCC 461

RESULT 3
US-10-085-783A-29177
; Sequence 29177, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29177
; LENGTH: 461
; TYPE: DNA
; ORGANISM: Human
US-10-085-783A-29177

Query Match 28.6%; Score 436.4; DB 7; Length 461;
Best Local Similarity 99.4%; Pred. No. 2.9e-96;
Matches 459; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 59 GGTGCCCATGGGAAGTCGGATTTCTTACTCCCAAGCTATCGCCACAGGATCAAGTC 118
Db 1 GGTGCCCATGGGAAGTCGGATTTCTTACTCCCAA-GCTATCGCCAAACAGGATCAAGTC 59
QY 119 CAAGGGCTGCAGAAAGCTACGCTGTTATTCAGAGATGTGCAGAAAGCAGTCCCGGACGA 178
Db 60 CAAGGGCTGCAGAAAGCTACGCTGTTATTCAGAGATGTGCAGAAAGCAGTCCCGGACGA 119
QY 179 GAATGGCTTTAAGTGTCAATGTATGTCGGAATCTCATCAGAGACAACATATTGCTGCTTC 238
Db 120 GAATGGCTTTAAGTGTCAATGTATGTCGGAATCTCATCAGAGACAACATATTGCTGCTTC 179
QY 239 AGAAATCCTCAGCAGTTTATGGAATATTTTTCAGAGAAATTCGAAATGACTTTCTAGA 298
Db 180 AGAAATCCTCAGCAGTTTATGGAATATTTTTCAGAGAAATTCGAAATGACTTTCTAGA 239
QY 299 ACTTCTCAGGAGACGCTTTGCGACTAAAGGGTCCACAACAACATTTCTTACACGAATA 358
Db 240 ACTTCTCAGGAGACGCTTTGCGACTAAAGGGTCCACAACAACATTTCTTACACGAATA 299
QY 359 CATCAGCCACCGAGAGACATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTT 418
Db 300 CATCAGCCACCGAGAGACATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTT 359
QY 419 TACTAAGTGTGCTGGC-AGAGAAAGCTTTGCAAAAGTGGACGAGACACCAAAAGGCTGGT 477
Db 360 TACTAAGTGTGCTGGC-AGAGAAAGCTTTGCAAAAGTGGACGAGACACCAAAAGGCTGGT 419
QY 478 ATATTCTAGTACATAGACAGGACCCAGAAACTATCCGCCGCC 519
Db ATATTCTAGTACATAGACAGGACCCAGAAACTATCCGCCGCC 461
```

Db 420 ATATTCACTACATAGACAGGACCCAGAACTATCCGCGGC 461

RESULT 4

US-11-097-143-41039
; Sequence 41039, Application US/11097143
; Publication No. US2005020858A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41039
; LENGTH: 1241
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-41039

Query Match 19.5%; Score 298.6; DB 10; Length 1241;
Best Local Similarity 54.2%; Pred. No. 3e-62;
Matches 638; Conservative 0; Mismatches 524; Indels 15; Gaps 1;

QY 64 CCATGGGAGTCGGATTTCTTACTCCCAAGCTATCGCCAAACAGATCAAGTCCAAG 123
DB 67 CGATGGTCGCGCGAGGTAGGTACGCCAAGTACCTCGCCAAACAGATCAAGTCCAAG 126

QY 124 GGCTGCAGAACTACGCTGTATTCAGATGTCCAGATGTCCAGAGCAGTCCCGGACGAGATG 193
DB 127 GTCTGCAGAACTCGCTGTATTCAGATGTCCAGATGTCCAGAGCAGTCCCGGATGAACAG 186

QY 184 GCTTTAAGTGTCAATGTATGTCCGAATCTCATCAGACAACTATTGCTGGCTTCAGAAA 243
DB 187 GCTTCAAGTGCACACATGAGCGAGTCCCAACAGCGCCAGTGTCTCTCTTTCGGGACA 246

QY 244 ATCCTCAGCAGTTTATGGATTTATTTTTCAGAGGAATTCGAAATGACTTTCTAGAACTTC 303
DB 247 ATCCTGCAAAATTCCTGCAGCTTCAGCAAGAGTTCCTCGACGGCTACATGGAGTTGC 306

QY 304 TCAGGAGACCTTTGGCACTAAAGGGTCCCAACACATGTCTACACGAATATACATCA 363
DB 307 TGCCTCGCGGCTTCGGCAGCAAGCGCAACCAAGTCTACAGAGGTACATTC 366

QY 364 GCCACCGAGAGCATCCATGAATGCCACTAGTGGGAACTCTGACTGATTTTACTA 423
DB 367 CCACAAGAGACATCCATGAACGCCACCCGATGGCTCACCTGTCCGACTACGTGA 426

QY 424 AGTGGCTGGGCGAGAAAGGCTTGTGCAAGTGGACGACACCAAAAAGGCTGGTATATTC 483
DB 427 AGTGGCTGGGCGGACTGGGCAAGTGTATCGGATGAGCGGAGAGGCTGGTTCGTCA 486
QY 484 AGTACATAGACAGGACCCAGAAACTATCCCGCGGCAACTGGAACTGGAGAAAAGAAA 543

Db 487 CCTACATTGATCGCAGTCCAGAGGCCATCGAAACGACAGCGGAGGCTGATCGCAAGGAGA 546
QY 544 AGCAGACCTTTCATGATGAAGGAAAAATGCGCAAAATTTATTCGAAGGCAAGTGCAGAGAG 603
DB 547 AGATGGAGAAAGGACGACGAGGAGCGGATGCGCACTTTCATTCAGCAGCAGATTTAAAAATG 606
QY 604 GCCTGGAAGGGAAGAAACAGGAGAGTCCCTATCTTTTACCGAAATTAAGCAGAGAAAAATGATG 663
DB 607 CCAAGGCCNAGGACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 666
QY 664 AAGAGAAAGTCACTGTTTAAATTTTGAAGTAAAGGAGCATGTAGCTCATCCGAGGACCAATCTT 723
DB 667 AAGAGAAACGAACCACTCAAGCTTGATATTCG-----CCTTGAGAAAAAGT 711
QY 724 CCAAGTCAAGTACTCTGGGACCGAGTGCACCTGAAGCAGATAGGAACTTCAGCATCATGTA 783
DB 712 TCCAGCCTGACACTGTGTAGGAAATCCGCTCTAGCCAGGACCTGCCCTTGAAGCTG 771
QY 784 AACGAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAGAAAGAAAGAAATCTG 843
DB 772 AAGAAAGGTGTTCAAGAAACCAAAATCCGTGGCTGGACACAGCCAAACCGCGTCGTCG 831
QY 844 CACTGATGAAATCATGAGATTTGAAGAGGAAAGAAAGAAAGTGCCTCCGAAACAGACTACT 903
DB 832 TGGACGAGATCATCAAGCAGGAGGAGGAAAGCAAAAGGAGGCTGCCAACCCGCAAGGACTACT 891
QY 904 GGCTACAGCTCAAAATTTTGAAGAAATTAACCAAGAAACTGGGAGAGAAATATCATATA 963
DB 892 GGCTGCACAAAGGCTATCGTGGTCAAAATTTTCCAAATTCATTCGGGCAAAAGTCTTCA 951
QY 964 AGAAAGGCTATTGTTAAGGAAGTAAATTCACAAATATACAGCTGTTGTAAGATGATG 1023
DB 952 ACAAAGGCGGTGTTCTGACGATTAATTCAGCATATCAGGCAAAATCAAGTCTTGG 1011
QY 1024 ATTTCTGGAGACAGCTGAAACTTGACCAGACTCATTTAGAGCAGATTAATTCAGACCCAG 1083
DB 1012 AGACTGGGAAAGCTAAAGTGATCAAGCTCATTTGGAGAGCGTAAATCCCGCTTTGG 1071
QY 1084 GAAAGAAATTCAGTTTAAATGGAGGCTACAGAGGAAATGAAGTACCTAGATCCCA 1143
DB 1072 ACAAGCTGTCTATGTTGTTAATGGCGCTTATCGGGGATCCGAGGCTCTGCTAAGGAAC 1131
QY 1144 TCAATCAGAGACTTTTTCAGCTACTATCTCATTTGAAACTGGCCCTTTTAAAGGAGCGCA 1203
DB 1132 TGGACGAGGCGCATATTCAGTCCGCTGGAATATTCAGCGTCTCTCAAGGCGAGAA 1191
QY 1204 GAGTTGAAGGAATTCATATGAAGACATTTCTAAACT 1240
DB 1192 TTGTAGACCAACGTCGAATACGAAGATATATCTAAACT 1228

RESULT 5
US-10-779-543-7646
; Sequence 7646, Application US/10779543
; Publication No. US20050227917A1
; GENERAL INFORMATION:
; APPLICANT: Williams et al
; TITLE OF INVENTION: GENE PRODUCTS DIFFERENTIALLY EXPRESSED
; TITLE OF INVENTION: IN CANCEROUS CELLS AND THEIR METHODS OF USE II
; FILE REFERENCE: 2300-21302
; CURRENT APPLICATION NUMBER: US/10/779,543
; CURRENT FILING DATE: 2004-02-12
; PRIOR APPLICATION NUMBER: 10/076,555
; PRIOR FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 09/217,471
; PRIOR FILING DATE: 1998-12-21
; PRIOR APPLICATION NUMBER: 60/068,755
; PRIOR FILING DATE: 1997-12-23
; PRIOR APPLICATION NUMBER: 60/080,664
; PRIOR FILING DATE: 1998-04-03
; PRIOR APPLICATION NUMBER: 60/105,234
; PRIOR FILING DATE: 1998-10-21

```
; PRIORITY NUMBER: 09/297,648
; PRIORITY FILING DATE: 2000-04-10
; PRIORITY APPLICATION NUMBER: PCT/US99/01619
; PRIORITY FILING DATE: 1999-01-28
; PRIORITY APPLICATION NUMBER: 60/072,910
; PRIORITY FILING DATE: 1998-01-28
; PRIORITY APPLICATION NUMBER: 60/075,954
; PRIORITY FILING DATE: 1998-02-24
; PRIORITY APPLICATION NUMBER: 60/080,114
; PRIORITY FILING DATE: 1998-03-31
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 23767
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7646
; LENGTH: 300
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-779-543-7646

Query Match      18.6%; Score 284; DB 9; Length 300;
Best Local Similarity 100.0%; Pred. No. 5.1e-59;
Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 634 CTTTACCGAATTAGCAGAGAAATGATGAGAGAAAGTCAAGTTTAAATTTGAGTAAAG 693
Db 17 CTTTACCGAATTAGCAGAGAAATGATGAGAGAAAGTCAAGTTTAAATTTGAGTAAAG 76

QY 694 GAGCATGTAGCTCATCCGGAGCAACATCTTCCAAGTCAAGTACTCTGGGACCGAGTGCAC 753
Db 77 GAGCATGTAGCTCATCCGGAGCAACATCTTCCAAGTCAAGTACTCTGGGACCGAGTGCAC 136

QY 754 TGAAGACATAGGAAGTTTCAGCATCAGTGAAGAAACGAAAGAAATCTTCCAGAGCTCAACTC 813
Db 137 TGAAGACATAGGAAGTTTCAGCATCAGTGAAGAAACGAAAGAAATCTTCCAGAGCTCAACTC 196

QY 814 AGTCTAAAGAAAGAAAGAAAGAAATCTGCACTCGATGAAATCATGAGATTGAAGAGG 873
Db 197 AGTCTAAAGAAAGAAAGAAAGAAATCTGCACTCGATGAAATCATGAGATTGAAGAGG 256

QY 874 AAAAGAAAGAACTGCCGAAACAGACTACTGGCTACAGCTGAA 917
Db 257 AAAAGAAAGAACTGCCGAAACAGACTACTGGCTACAGCTGAA 300

RESULT 6
US-11-097-143-41038
; Sequence 41038, Application US/11097143
; Publication No. US2005020858A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; PRIORITY FILING DATE: 2005-04-04
; PRIORITY APPLICATION NUMBER: 60/157,832
; PRIORITY FILING DATE: 1999-10-05
; PRIORITY APPLICATION NUMBER: 60/160,191
; PRIORITY FILING DATE: 1999-10-19
; PRIORITY APPLICATION NUMBER: 60/161,932
; PRIORITY FILING DATE: 1999-10-28
; PRIORITY APPLICATION NUMBER: 60/164,769
; PRIORITY FILING DATE: 1999-11-12
; PRIORITY APPLICATION NUMBER: 60/173,383
; PRIORITY FILING DATE: 1999-12-28
; PRIORITY APPLICATION NUMBER: 60/175,693
; PRIORITY FILING DATE: 2000-01-12
; PRIORITY APPLICATION NUMBER: 60/184,831
; PRIORITY FILING DATE: 2000-02-24
; PRIORITY APPLICATION NUMBER: 60/191,637
; PRIORITY FILING DATE: 2000-03-23

; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41038
; LENGTH: 3313
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-41038

Query Match      16.2%; Score 247.2; DB 10; Length 3313;
Best Local Similarity 54.1%; Pred. No. 2e-49;
Matches 535; Conservative 0; Mismatches 438; Indels 15; Gaps 1;

QY 64 CCATGGGGAAGTCGGATTTCTTACTCCCAAGGCTATCGCCAAAGAGTCAAGTCCAGG 123
Db 1087 CGATGGGTGCGCGGAGGTAGGTACGCCCAAGTACCTCGCCAAAGATGAATCGAAGG 1126

QY 124 GGCTGCAGAAAGCTACGCTGGTATTGCCAGATGTGCCAGAGCAGTCCCGGACGAGAATG 183
Db 1127 GTCTGCAGAAAGCTGGCTGGTACTGCCAGATGTGGGAGAGCAGTCCCGATGAACG 1186

QY 184 GCTTTAAGTGTCAATTGTATGTCGGAATCTCATAGAGACAACTATTTGTGCTTCAGAAA 243
Db 1187 GCTTCAAGTGCACACGATGAGCGAGTCCACACGCGCAGTTGCTCTCTTTGGGACA 1246

QY 244 ATCTTCAGCAGTTTATGATTTTTCAGAGGAATTCGGAATGACTTTCTAGAACTTC 303
Db 1247 ATCTTGGCAAAATTCCTGCACAGCTTCAGCAAGAGTTCTCCGACGGCTACATGGAGTTGC 1306

QY 304 TCAGGAGACGCTTTGGGCACTAAAAGGTTCCAAACCAACATTTGTCTACAAAGAAATACATCA 363
Db 1307 TGGCGCGGCTTGGGACGAGCGAACCGAGCGCCACAGATCTACGAGGATACATTG 1366

QY 364 GCCACCGAGAGACATCCACATGAATGCCATCAGTGGGAAACTCTGACTGATTTTACTA 423
Db 1367 CCCACAGGAGACATCCACATGAACGCCACCGATGCTCACTCCGCTCCGACTACGTGA 1426

QY 424 AGTGGCTGGGAGAGAGGCTTGTGCAAGTGGAGCAGACACCAAGAAAGCTGGTATATTC 483
Db 1427 AGTGGCTGGGCGGACTGGGCAAGTGATAGCGGATGAGACGGAGAGGCTGGTTCGTCA 1486

QY 484 AGTACATAGACAGGAGACCCAGAAACTATCCCGCGCAACTGGAACTGGAGAAAAAGAAAA 543
Db 1487 CCTACATTGATCGCAGTCCAGAGGCCATGGAAACGACAGCGGAGGCTGATCGCAAGGAGA 1546

QY 544 AGCAGGACCTTTGATGATGAAGAAAAAACTGCCAAATTTATTGAAGAGCAAGTGAAGAGAG 603
Db 1547 AGATGGAGAGAGACGACGAGGAGCGGATGGCGGCTTCATTTCAGCAGCAGATTAAGAAATG 1606

QY 604 GCCTGGAAGGGAAGGACGAGGAGGTCCTCTCTTTTACGGAATTAAGCAGAGAAAAATGATG 663
Db 1607 CCAAGGCCAAGGAGCGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1666

QY 664 AAGAGAAAGTCAAGCTTTTAAATTTGAGTAAAGGAGCATGTAGCTCATCCGAGCAACATCTT 723
Db 1667 AAGAGAACGAACCACTCAAGCTTGTATTCG-----CCTTGAGAAAAAGT 1711

QY 724 CCAAGTCAAGTACTCTGGGACCGAGTGCACTGAAGACGATAGGAAAGTTCAGCATCAGTGA 783
Db 1712 TCCAGCCTGACACTGTCTAGGGAATCCGCTCTAGCCAAGCGACCTGCCCTGAGGCTG 1771

QY 784 AACGAAAGAAATCTTCCAGAGCTCAACTCAGTCTTAAAGAAAAAGAAAAAGAAATCTG 843
Db 1772 AAGAAAGAGTGTCAAGAAACCCCAATCCGTGGCTGGAGACAGCCAAACGCGGTCGGTGC 1831

QY 844 CACTGGATGAATCATGGAGATTGAAGAGGAAAAAGAAAGAAAGTCCCGCAACAGACTACT 903
Db 1832 TGGACGAGATCATCAGCAGGAGGAAAGCAAAAGAGGCGGTGCCAACCCGAGGACTACT 1891

QY 904 GGCTACAGCCTGAAATTTATTGTGAAAAATTAACCAAGAAACTGGGAGAGAAATATCAT 963
Db 1892 GGCTGCAACAGGATTCGTGGTCAAAATTTATTTCAAATCCATCCATGGGCGAAAAAGTTCTTCA 1951

QY 964 AGAAAAAGGCTATTGTTAAGGAAGTAAATTTGACAAATATACAGCTGTTGTTGAGAGATGTTG 1023
```

Db 1952 AACAAAAAGCGTGTCTGGACGTAATGACAGATATCAGGCGCAAAATCAAGTCTTGG 2011
Qy 1024 ATTCTGGAGACAAAGCTGAAACTTGACCA 1051
Db 2012 AGACTGGGGAAGGCTAAAAAGTGGATCA 2039
RESULT 7
US-11-097-143-41017
; Sequence 41017, Application US/11097143
; Publication No. US2005020858A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41017
; LENGTH: 4394
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-41017
Query Match 13.7%; Score 209.4; DB 10; Length 4394;
Best Local Similarity 58.6%; Pred. No. 4.3e-40;
Matches 363; Conservative 0; Mismatches 256; Indels 0; Gaps 0;
Qy 64 CCATGGGGAAGTCGGATTTCTTACTCCCAAGGCTATCGCCAAACAGATCAAGTCCAAGG 123
Db 3756 CGATGGTTCGCGCGAGGTAGGTACGCCCAAGTACCTCGCCAAACAGATGAATCGAAGG 3815
Qy 124 GGCTGCAGAGCTACGCTGTATTGTCAGATGTCAGAGAGCAGTGC CGGACGAGAATG 183
Db 3816 GTCTGCAAGAGCTGCGCTGTGATCTGCCAGATGTCGAGAGCAGTGC CGGATGAACG 3875
Qy 184 GCTTTAAGTGTCAATGATGTCGGAATCTCATCAGAGACAATTAATGCTGGCTTCAGAAA 243
Db 3876 GCTTCAAGTCCACACGATGACGAGTCCCAACAGCGCAGTGTGCTCTCTTTGCGGACA 3935
Qy 244 ATCTCAGCAGTTATGGATTAATTTTCAGAGGAAATTCGGAATGACTTTCTAGAACTTC 303
Db 3936 ATCTGCAAAATCTCTCAGCAGCTTCAGCAAGAGTTCCTCCGACGGCTACATGGAGTTGC 3995
Qy 304 TCAGGAGACGTTTGGCACTTAAAGGTCACAAACACATTGTCTACAAAGATATACATCA 363
Db 3996 TGGCCGCGCGGTTCGGCACAAGAGGAAACCAACGATCTACAGAGTACATTTG 4055
Qy 364 GCCACCCAGAGACATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTTTACTA 423
Db 4056 CCACAAAGAGACATCCACATGACGCCACCCGATGGCTCACCTGTGCTGACTACGTGA 4115

Qy 424 AGTGGCTGGCAGAGAGGCTTTGTGCAAAAGTGGACGAGACACCAAAAGGCTGGTATATTC 483
Db 4116 AGTGGCTGGCAGGACTGGCAAGTGATAGCGGATGAGACGGAAGGCTGTTCTGTCGA 4175
Qy 484 AGTACATAGACAGGACCCAGAAACTATTCGCGCGCACTGGAACCTGGAGAAAAAGAAA 543
Db 4176 CCTACATTGATCGCAAGTCCAGAGGCCATGGAACGACAGCGGAGGCTGATCCCAAGGAG 4235
Qy 544 AGCAGGACCTTGATGATGAAGAAAAAACTGCCAAAATTTATTGAAGAGCAAGTGAGAAGAG 603
Db 4236 AGATGAGAGAGACGACGAGGCGGATGCGCGACTTCATTGAGCAGCAGATTAATAATG 4295
Qy 604 GCCTGGAAGGGAAGAACAGAGGAGTCCCTACTTTTTCGGAATTAAGCAGAGAAAAATGATG 663
Db 4296 CCAAGGCCAAGGACGCGGAGGAGGACGAAGGCCAGAGAAGTTTACCGAGCTAAAGCGCG 4355
Qy 664 AAGAGAAAGTCACTTTAA 682
Db 4356 AAGAGAAAGCAACCACTCA 4374
RESULT 8
US-11-097-143-24532
; Sequence 24532, Application US/11097143
; Publication No. US2005020858A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24532
; LENGTH: 4582
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-24532
Query Match 13.7%; Score 209.4; DB 10; Length 4582;
Best Local Similarity 58.6%; Pred. No. 4.4e-40;
Matches 363; Conservative 0; Mismatches 256; Indels 0; Gaps 0;
Qy 64 CCATGGGGAAGTCGGATTTCTTACTCCCAAGGCTATCGCCAAACAGATCAAGTCCAAGG 123
Db 3944 CGATGGTTCGCGCGAGGTAGGTACGCCCAAGTACCTCGCCAAACAGATGAATCGAAGG 4003
Qy 124 GGCTGCAGAGCTACGCTGTATTGTCAGATGTCGAGAGCAGTGC CGGACGAGAATG 183
Db 4004 GTCTGCAAGAGCTGCGCTGTGATCTGCCAGATGTCGAGAGCAGTGC CGGATGAACG 4063
Qy 184 GCTTTAAGTGTCAATGATGTCGGAATCTCATCAGAGACAATTAATGCTGGCTTCAGAAA 243
Db 4064 GCTTCAAGTGCACACGATGAGCGAGTCCCAACGAGGCTGCTCTCTCTTTGCGGACA 4123

QY 244 ATCCCTCAGAGTTTATGATTTATTTTCAGAGGAAATTCGAAATGACTTTCTAGAACTTC 303
DB 4124 ATCCCTGGCAAAATTCCTGCACAGCTTCAGCAAAAGAGTTCTCCGACGGCTACATGGAGTTGC 4183
QY 304 TCAGGAGAGCGTTTCGCACTAAAGGGTCCACAAACAATTTGTCTACAAAGAAATACATCA 363
DB 4184 TCGCGCGGGTTCGGCAGAAAGCGAACCCAGCGCAACAAGATCTACCAAGGAGTACATTG 4243
QY 364 GCCACCGAGAGCATTCACATGAATGCGACTTCAGTGGGAAACTCTGACTGATTTTACTTA 423
DB 4244 CCCACAAGGAGCACATCCACATGAACGCCACCGGATGGCTCACTGTCGACTACGTGA 4303
QY 424 AGTGCTGGCAGAGAAAGGCTTGTGCAAGTGGAGAGACACCAAAAGGCTGGTATATTC 483
DB 4304 AGTGGCTGGCGGAGCTGGGCAAGTGATAGCGGATGAGACGAGAGGGGCTGGTTCGTCA 4363
QY 484 AGTACATAGACAGGAGCCAGAAACTATCCGCGCGCACTGGAACTGGAGAAAGAGAAA 543
DB 4364 CCTACATTGATCGAGTCCAGAGGCCATGGAAACGACAGCGGAAGGCTGATGCCAAGGAGA 4423
QY 544 AGCAGGACCTTGATGATGAAGAAAAAACTGCCAAATTTATTTGAAGAGCAAGTGAAGAG 603
DB 4424 AGATCGAAGAGACGACGAGGAGCGATGGCCGACTTCATTGAGCAGCAGATTAABAATG 4483
QY 604 GCCTGGAAGGGAAGGAACAGAGAGGTCCTTACTTTTACGGAATTAAGCAGAGAAAAATGATG 663
DB 4484 CCAAGGCCAAGACGCGGAGGAGGACGAAGGCCAGGAGAAGTTTCCCGAGCTTAAAGCGCG 4543
QY 664 AAGAGAAATGTCAGTTTAA 682
DB 4544 AAGAGAACGAACCACTCA 4562

RESULT 9

US-10-424-599-25484
; Sequence 25484, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 25484
; LENGTH: 777
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MBT3847_123013C.1
US-10-424-599-25484

Query Match 13.5%; Score 205.6; DB 7; Length 777;
Best Local Similarity 62.8%; Pred. No. 1.4e-39;
Matches 319; Conservative 0; Mismatches 189; Indels 0; Gaps 0;
QY 61 TCGCCATGGGGAAGTCGGATTTTCTTACTCCCAAGGCTATCCCAACAGGATCAAGTCA 120
DB 135 TCGTGTAGGGGAAAAATGAGTTTCTCACACCTTAAAGCAATTCCTCAATCAAAAGCAA 194
QY 121 AGGGCTCGAAGACTACGCTGATTCGACATGTCAGAGAGAGTCCCGGAGCAGA 180
DB 195 AAGGATTCGAGAAGCTTCGGTGGTATTCGACATGTGTGAGAGAGAGTCCCGAGATGAGA 254
QY 181 ATGGCTTTAAAGTGTCAATGTATGTCCGAATCTCATCAGAGACAACTATTGCTGGCTTCAG 240
DB 255 ATGGCTTTAAATGCCATTGCATGAGTGAAGGCCACCAAGCGTCAATGCNAGATTTTGGAC 314

QY 241 AAAATCCTCAGACGTTTATGATTTATTTTCAGAGGAAATTCGAAATGACTTTCTAGAAC 300
DB 315 AAAACCCACACCGGATAGTTGAGGGCTATTCCGAAGAGATTGAGAGTACTTTTCTGGAGC 374
QY 301 TTCTCAGAGAGCGCTTTGGCACTAAAAGGGTCCACAAACAATTTGTCTACAAAGAAATACA 360
DB 375 ACATGAAGCGAGTCACCGATTACCGGTGTGGCAGCCACTGTAGTTTATACGAAATACA 434
QY 361 TCAGCCACCGAGAGACATCCACATGAATGCCATCTAGTGGAAACTCTGACTGATTTTA 420
DB 435 TAAATGACAGACACCAACATTATGAATCTACTCAGTGGGCTACGCTTACTGAGTTTG 494
QY 421 CTAAGTGGCTGGCAGAGAGGCTTGTGCAAAAGTGGACGACACCAAAAGGCTGGTATA 480
DB 495 TTAAGTACTTGGGTGCGAACTGGCAAAATGTAAAGTTGAGGAAACACCCAAAGGATGGTTCA 554
QY 481 TTCAAGTACATAGACAGGACCCAGAACTATCCGCGCGCAACTGGAACTGGAGAAAAAGA 540
DB 555 TTACATATATAGATAGAGATTTCAGAAACCTTTTTCAGGAGAGGATGAAGATAAGAGAA 614
QY 541 AAAAGCAGGACCTTGATGATGAAGAAAA 568
DB 615 TCAAGGCAGATATGTTAGTAGAAGAAAA 642

RESULT 10

US-10-739-930-2955
; Sequence 2955, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 2955
; LENGTH: 1713
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: ZEAMA-23APR03-CLUSTER7583_2
US-10-739-930-2955

Query Match 13.4%; Score 204.8; DB 8; Length 1713;
Best Local Similarity 62.9%; Pred. No. 3.4e-39;
Matches 317; Conservative 0; Mismatches 187; Indels 0; Gaps 0;
QY 62 CGCCATGGGGAAGTCGGATTTTCTTACTCCCAAGGCTATCCCAACAGGATCAAGTCCAA 121
DB 171 CGCGATGGGGAAGCAGCAGAGTTCTTGACGCCGGAAGCGATCCGAAACCGAGCAAGGCCAA 230
QY 122 GGGCTGCAGAGCTACGCTGATTCGACAGATGTCAGAGAGAGTCCCGGAGCAGAA 181
DB 231 GGGCTGCAGAGCTGCGGTGGTACTGCCAGATGTGTGAGAGAGAGTCCCGGAGCAGAA 290
QY 182 TGGCTTTAAGTGTCAATGTATGTCCGAATCTCATCAGAGACAACTATTGCTGGCTTCAGA 241
DB 291 CGGGTTCAAGTGCCACTGCATGTCCGAGTCGACACCGGCGAGATGCAAGGTTCGCGCAT 350
QY 242 AAATCCTCAGCAGTTTATGGATTTATTTTCAGAGAAATTCGAAATGACTTTCTAGAACT 301
DB 351 GCGCGCCGACCGCGTCGTCGAGGGCTTCTCCGAGGAAATTCCTCGAGTCTCTCTCCCT 410
QY 302 TCTCAGGAGACGCTTTGGCACTAAAAGGTCACAAACAATTTGTCTACAAAGAAATACAT 361
DB 411 CATCCGCGCGGACCGCCACTCCCGGTGCGCGCCACCGCTTGTCTACAAAGAGTACAT 470
QY 362 CAGCCACCGGAGCAGACATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTTTAC 421
DB 471 CGCGCAGCCGACCAACCGCTCCACATGAACTCTACGCGGTGGGCCCAAGCTCACCGAGTTCT 530

QY 422 TAAGTGGCTGGCAGAGAGGCTTGTGCAAACTGGACGAGACCAAAAGGCTGTATAT 481
DB 531 CAAGCTCTCTGGGGCGGGAAGGGTACTGCAAGGTTGAGGACGCCCAAGGGGTGTTCTAT 590
QY 482 TCAGTACATAGACAGGAGCCCAAGAACTATCCGCGGCAACTGGAACTGGAGAAAGAA 541
DB 591 GACCTACATCGACGCTGACTCGGAGCAGGCCCTCAAGGACCGCTCAAGCGCAGAGGAT 650
QY 542 AAAGCAGGACCTTGATGATGAAGA 565
DB 651 CAAGTCAGACATGGCTGAGGACGA 674

RESULT 11

US-10-425-115-157125
; Sequence 157125, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5322)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 157125
; LENGTH: 2046
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_74879C.1
US-10-425-115-157125

Query Match 13.3%; Score 203.2; DB 8; Length 2046;
Best Local Similarity 62.7%; Pred. No. 9.3e-39;
Matches 316; Conservative 0; Mismatches 188; Indels 0; Gaps 0;
QY 62 CGCCATGGGGAAGTCGGATTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCAA 121
DB 294 CGCGATGGGGAAGACAGAGTTCTGAGCGCGAAGGCGATCGCAATCGGATCAAGGCGAA 353
QY 122 GGGGCTGCAGAGCTACGCTGGTATTGCCAGATGTGCCAAGAGAGTGGCGGAGCAGAA 181
DB 354 GGGGCTGCAGAGCTGCGGTGGTACTGCCAGATGTGTCAAGACAGTGGCGGACGAGAA 413
QY 182 TGGCTTTAAGTGTCAATGTATGTCGNAATCTCATCAGAGACACTATTGCTGGCTTCAGA 241
DB 414 CGGGTTCAAGTGCACCTGCAATGTGGAGTGCACACGAGCAGATGCAGGTGTTGCGCAT 473
QY 242 AATCTCTCAGCAGTTTATGATGATTTTTCAGAGGAATTCGAAATGACTTTCTAGAACT 301
DB 474 GGGCCCGCAGCCGCTGCTCGAGGGCTTCTCGAGGAATTTCTCGAGTCTTCTCTCCCT 533
QY 302 TCTCAGGAGACGCTTTGGCACTAAAAGGGTCCACAAACATTTGTCTAACAGGAATACAT 361
DB 534 CATCCGCGCGCGCAGCCGCACTCCCGCTCGCGCCACCGTTGTCTTACACAGGATACAT 593
QY 362 CAGCCACGAGAGACATCCACATGAATGCCACTCAGTGGGAACCTGTGACTGATTTTAC 421
DB 594 CGCCGACCGGACCAACGCTCCACATGAATCTTACGCGGTGGGCGCACCGCTCACCGAGTTCTG 653
QY 422 TAAGTGGCTGGGAGAGAGGCTTGTGCAAAAGTGGACGACACCAAAAGGCTGGTATAT 481
DB 654 CAACTCTGGGGCGCGAAGGGTACTGCAAGGTTGAGGACACGCCCAAGGGGTGTTCTAT 713
QY 482 TCAGTACATAGACAGGAGCCCAAGAACTATCCGCGGCAACTGGAACCTGGAGAAAGAA 541
DB 714 GACCTACATCGACCTGACTCGGAGCAGGCGCTCAAGGACCGCTCAAGCGCAGAGGAT 773
QY 542 AAAGCAGGACCTTGATGATGAAGA 565

RESULT 13
US-10-425-114-22614
; Sequence 22614, Application US/10425114

DB 774 CAAGTCAGACATGGCTGACGACGA 797
RESULT 12
US-10-425-114-2975
; Sequence 2975, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 2975
; LENGTH: 1614
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700238628_FLI
US-10-425-114-2975

Query Match 13.2%; Score 201.6; DB 7; Length 1614;
Best Local Similarity 62.5%; Pred. No. 2e-38;
Matches 315; Conservative 0; Mismatches 189; Indels 0; Gaps 0;
QY 62 CGCCATGGGGAAGTCGGATTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCAA 121
DB 76 CACGATGGGGAAGACAGAGTTCTGACGCGGAAGGCGATCGGAAACCGGATCAAGGCGAA 135
QY 122 GGGGCTGCAGAGCTACGCTGGTATTGCCAGATGTGCCAAGAGCAGTGGCGGAGCAGAA 181
DB 136 GGGGCTGCAGAGCTGCGGTGGTACTGTCTAGATGTGTCTAGAGCAGTGGCGGACGAGAA 195
QY 182 TGGCTTTAAGTGTCAATGTATGTCGNAATCTCATCAGAGACACTATTGCTGGCTTCAGA 241
DB 196 CGGGTTCAAGTGCACCTGCAATGTGCGAGTGCACACGAGCAGATGCAGGTGTTGCGCAT 255
QY 242 AATCTCTCAGCAGTTTATGATGATTTTTCAGAGGAATTCGAAATGACTTTCTAGAACT 301
DB 256 GGGCCCGCAGCCGCTGCTCGAGGGCTTCTCCGAGGAGTTCTCGAGTCTTCTCTCCCT 315
QY 302 TCTCAGGAGACGCTTTGGCACTAAAAGGGTCCACAAACATTTGTCTTACAAAGAAATACAT 361
DB 316 CATCCGCGCGGCGCACCGCACTCCCGCTGCGCGCCACCGTCTGTCTACAAAGGATACAT 375
QY 362 CAGCCACGAGAGACATCCACATGAATGCCACTCAGTGGGAACTCTGTGATTTTAC 421
DB 376 CGCCGACCGGACCACTGTCACATGAATCCACGCGTGGGCGCACCGCTCAACCGAGTTCTG 435
QY 422 TAAGTGGCTGGGAGAGAGGCTTGTGCAAAAGTGGAGCAGACACCAAAAGGCTGGTATAT 481
DB 436 CAAAGTTCTTGGGCGGCGAGGGGTACTGTAAAGTTCGAGGACACGCCCAAGGGGTGTTCTAT 495
QY 482 TCAGTACATAGACAGGAGCCCAAGAACTATCCGCGGCAACTGGAACCTGGAGAAAGAA 541
DB 496 GACCTACATCGACCTGATCTCGGAGCAGGCGCTCAAGGACCGCTCAAGCGCAGAGGAT 555
QY 542 AAAGCAGGACCTTGATGATGAAGA 565
DB 556 CAAGTCAGACATGGCTGACGACGA 579

```

; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(531313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 22614
; LENGTH: 1713
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3591-011-E3_FLI
US-10-425-114-22614

Query Match      13.2%; Score 201.6; DB 7; Length 1713;
Best Local Similarity 62.5%; Pred. No. 2.1e-38;
Matches 315; Conservative 0; Mismatches 189; Indels 0; Gaps 0;

QY      62  CGCCATGGGGAAGTCGGATTTCCTTACTCCCAAGGCTATCGCAACAGAGATCAAGTCCAA 121
Db      235  CACATGGGGAAGCACGAGTTCCTGACGCCGAGCGGATCGCAACCGGATCAAGCGNA 294

QY      122  GGGGCTGCAGAAGCTACGCTGGTATTGCCAGATGTGCCAAGACGAGTGC CGGACGAGAA 181
Db      295  GGGGCTGCAGAAGCTGCGGTGGTACTGTGAGATGTTCAGAAGCAGTGTCCGCGACGAGAA 354

QY      182  TGGCTTTAAGTGTGATTGTATGTCCGAATCTCATCAGAGACAACTATTGCTGGCTTCAGA 241
Db      355  CGGGTTCAAGTGCCACTGTCATGTCCGAGTGCACACGAGGCGAGATGCGAGTTC CGGCAT 414

QY      242  AATCTCTCAGCAGTTATGGAATTATTTTCAGAGGAATTCGAAATGACTTTCTAGAACT 301
Db      415  GGCCCCGACCGGTGTCGAGGGCTTCTCCGAGAGTTCCTCGAGTCTCTCTCTCCCT 474

QY      302  TCTCAGGAGACGCTTTGSCACTAAAGGTTCCACAACAACATTTGTCTACAACGAATACAT 361
Db      475  CATCGCGCGCGCACCGCCACTCCCGCGTCGCCGCCACCGTCGTCTACAACGAGTACAT 534

QY      362  CAGCCACGAGAGACATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTTTAC 421
Db      535  CGCCGACCGGCACCAACGCTCCACATGAATCCACGCGTGGGCGACGCTCACCAGTTCTGT 594

QY      422  TAAGTGGCTGGGACAGAAAGCTTGTGCAAGTGGACGAGACACCAAAAGGCTGTATAT 481
Db      595  CAAGTTCCTGGGGCGGAGGGGTACTGTAAAGTCGAGGACACGCCCAAGGGGTGTTTAT 654

QY      482  TCAGTACATAGACAGGACCCAGAAACTATCCGCCGGCAACTGGAACCTGGAAAAAGAA 541
Db      655  GACCTACATCGACCGTGACTCGGAGCAGCGCTCAAGGNAAGCGCTCAAGCGCAAGAGGAT 714

QY      542  AAAGCAGGACCTTGATGATGAAGA 565
Db      715  CAAGTCAGACATGGCTGACGACGA 738

```

; APPLICANT: Boukharov, Andrey A.
 ; APPLICANT: Barbazuk, Brad
 ; APPLICANT: Li, Ping
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53221)B
 ; CURRENT APPLICATION NUMBER: US/10/437,963
 ; CURRENT FILING DATE: 2003-05-14
 ; NUMBER OF SEQ ID NOS: 204966
 ; SEQ ID NO 89433
 ; LENGTH: 538
 ; TYPE: DNA
 ; ORGANISM: Oryza sativa
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT4530_88190C.1
 US-10-437-963-89433

Query Match 12.6%; Score 192; DB 7; Length 538;
 Best Local Similarity 65.8%; Pred. No. 2.4e-36;
 Matches 279; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

Qy	54	GCGGTGTCGCGATGGGGAAGTCGAGATTCTTACTCCCAAGGCTATGCCCAACAGGATC	113
Db	112	GTCGGCAAGAGATGGGGAACACAGATTCTGACGCCGAGGCGATCCGGAACAGGATC	171
Qy	114	AAGTCCAAAGGGGCTGCAGAAAGCTACGCTGGTATTGCCAGATGTGCCAAGCAGTGC	173
Db	172	AAGCGAAGGGCTGCAGAAAGCTGCGGTGGTACTGCCAGATGTGCCAAGCAGTGC	231
Qy	174	GACGAGAAATGGCTTTAAGTGTATGTATGTCGGAATCTCATCAGAGACAACTATTGCTG	233
Db	232	GACGAGAAATGGCTTTCAAGTGCACCTGCAATGTGCGAGTGCACACGCGCCAGATGCAGGTG	291
Qy	234	GTTCAGAAATCTCAGCAGTTTATGGATTATTTTCAGAGGAATTCGAAATGACTTT	293
Db	292	TTCGGCCAGGCCCCCGAGTCGTCGAGGGCTTCTCCGAGGAGTTCTCTGACGCTTC	351
Qy	294	CTAGAACTTCTCAGGAGACGCTTTGGGCACATAAAGGGTCCCAACAACTTCTTACAAAC	353
Db	352	CTACCTTGTCTCCGCGGCCACCGACACTCCCGCATCGCGCCACCGTCTGTACAAAC	411
Qy	354	GAATACATCAGCCACGAGAGACATCCACATGAATGCCACTCAGTGGGAAACTCTGACT	413
Db	412	GAGTTTCATCGCGACCGCCACACCGTCCACATGAATCCACGCGCTGGGCCACGCTCAC	471
Qy	414	GATTTTACTAAGTGGCTGGGCGAGAGAGCTTGTGCAAGTCGAGAGACACCAAAAGGC	473
Db	472	GAGTTTGTCAAGTTCTCTGGGCGCGAGGGCCACTGCAAGTTGAGGACACTCCCAAGGG	531
Qy	474	TGGT 477	
Db	532	TGGT 535	

RESULT 15
 US-10-767-701-25587
 ; Sequence 25587, Application US/10767701
 ; Publication No. US20040172684N1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
 ; FILE REFERENCE: 38-21(53535)B
 ; CURRENT APPLICATION NUMBER: US/10/767,701
 ; CURRENT FILING DATE: 2004-01-29
 ; NUMBER OF SEQ ID NOS: 63128
 ; SEQ ID NO 25587
 ; LENGTH: 584
 ; TYPE: DNA
 ; ORGANISM: Sorghum bicolor
 ; FEATURE:

OTHER INFORMATION: Clone ID: 30963818
US-10-767-701-25587

Query Match 10.3%; Score 156.8; DB 7; Length 584;
.Best Local Similarity 64.7%; Pred.No.1.1e-27;
Matches 233; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

Qy	66	ATGGGGAAGTCGGATTCTTCTACTCCRAAGGCTATCGCCACAGGATCAAGTCCAAGGGG	125
Db	225	ATGGGGAAGTCGGATTCTTCTAGCGCGAAGGGGATCGCGAACCGGATCAAGGCAAGGGG	284
Qy	126	CTGCAGAACTACGCTGGTATTGCCAGATGTGCCAGAGCAGTGC CGGACGAGAAATGGC	185
Db	285	CTGCAGAAAGCTGGGTGGTACTGCCAGATGTGCCAGAGCAGTGC CGGACGAGAAATGGC	344
Qy	186	TTTAAGTGTCAATGTATGTCCGAATCTCATCAGAGACAACTATTGCTGGCTTCAGAAAT	245
Db	345	TTCAAGTGCACACTGCATGTCCGAGTCGCACCGCGCAGATGCAGGTGTTCCGCCATGGCG	404
Qy	246	CCTCAGCAGTTTATGGATTATTTTCAGAGGAATTCGGAATGACTTTCTAGAACTTCTC	305
Db	405	CCCGACCGCGTCGTCGAGGGCTTCTCCGAGGAGTTCTCGAGTCTCTCTCCCTCATC	464
Qy	306	AGGAGACGCTTTGGCACTAAAGGGTCCACAAACATTTGTCTACAAAGAAATACATCAGC	365
Db	465	CGCGCGCGGACCGCCACTCCCGCGTCCGCGCCACCGTCTCTACAAAGGATACATCGCG	524
Qy	366	CACCGAGAGACATCCACATGAATGCCACTCAGTGGGAACTCTGACTGATTTTACTAAG	425
Db	525	GACCGTCAACCACTCCACATGAACTCCAGCGCTGGGCGCACGCTCACCAGATTCTGTCAG	584

Search completed: November 27, 2005, 02:22:24
Job time : 1765.53 secs

This Page Blank (uspdc)

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	44	2.9	437	1	US-10-021-234-402	Sequence 402, App	
2	40.4	2.6	1575	7	US-11-074-176-145	Sequence 145, App	
3	39.4	2.6	3254	1	US-10-793-626-4202	Sequence 4202, App	
C 4	39	2.6	2970	1	US-10-793-626-2593	Sequence 2593, App	
	39	2.6	3549	1	US-10-793-626-3796	Sequence 3796, App	
C 6	37.8	2.5	966	1	US-10-793-626-1525	Sequence 1525, App	
	37.8	2.5	3742	1	US-10-793-626-4076	Sequence 4076, App	
7	37.8	2.5	3782	1	US-10-793-626-3753	Sequence 3753, App	
8	37.8	2.5	3985	1	US-10-793-626-3856	Sequence 3856, App	
9	37.2	2.4	857	9	US-11-033-764-56	Sequence 56, Appl	
10	37.2	2.4	3927	1	US-10-793-626-4223	Sequence 4223, App	
11	37.2	2.4	4189	1	US-10-793-626-3898	Sequence 3898, App	
12	37.2	2.4	4800	7	US-11-159-597-20	Sequence 20, Appl	
13	37.2	2.4	4000	7	US-10-793-626-2251	Sequence 2251, App	
14	37	2.4	345	1	US-10-793-626-3415	Sequence 3415, App	
15	37	2.4	3153	1	US-10-793-626-3415	Sequence 3415, App	
C 16	37	2.4	3555	1	US-10-793-626-3996	Sequence 3996, App	
	37	2.4	3656	1	US-10-793-626-3543	Sequence 3543, App	
17	36	2.4	34000	7	US-11-102-978-3	Sequence 3, Appl	
18	36	2.4	34000	7	US-11-102-978-3	Sequence 3, Appl	
C 19	35.8	2.3	2913	1	US-10-793-626-3464	Sequence 3464, App	
	35.8	2.3	4017	1	US-10-793-626-4104	Sequence 4104, App	
C 20	35.8	2.3	3041	1	US-10-793-626-4368	Sequence 4368, App	
C 21	35.6	2.3	3041	1	US-10-793-626-4368	Sequence 4368, App	
C 22	35.6	2.3	3509	9	US-11-077-550-19	Sequence 19, Appl	
C 23	35.6	2.3	34000	7	US-11-102-978-3	Sequence 3, Appl	

RESULT 2

US-11-074-176-145
; Sequence 145, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
; APPLICANT: Kluenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Peril, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; TITLE OF INVENTION: Stress-Related Proteins and Uses Therefore
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074,176
; PRIOR FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: 60/551,161
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 145
; LENGTH: 1575
; TYPE: DNA
; ORGANISM: Lactobacillus acidophilus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1575)
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: ORF 1524; Lisk - Putative histidine kinase/may be
; OTHER INFORMATION: involved in stress response
US-11-074-176-145

Query Match 2.6%; Score 40.4; DB 7; Length 1575;
Best Local Similarity 53.9%; Pred. No. 0.57; Mismatches 71; Indels 0; Gaps 0;
Matches 83; Conservative 0;
QY 915 GAAATATTGTGAAATTAACCAAGAACTGGGAGAGAAATATCATAAGAAAAAGGCT 974
DB 1003 GAAATGCTTGATTAACTAGAGCTGAACAAATTGATGTTCAATATCCITATGAAGTAACA 1062
QY 975 ATTGTTAAGGAAGTAATTGACAAATATACAGCTGTTGTGAAGATGATTCTCGAGAC 1034
DB 1063 AATGTTAAGCAACAGTAAACAGAGTTGTTCTGATTTGGCAATGTTTCATTTCAGACTTT 1122
QY 1035 AAGCTGAACCTTGACCAGACTCATTTAGAGACAG 1068
DB 1123 AAGATTCAACTTGATGAAGATGATTTTACCACCAG 1156

RESULT 3

US-10-793-626-4202
; Sequence 4202, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4202
; LENGTH: 3254
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence

US-10-793-626-4202

Query Match 2.6%; Score 39.4; DB 1; Length 3254;
Best Local Similarity 47.7%; Pred. No. 1.5;
Matches 115; Conservative 0; Mismatches 126; Indels 0; Gaps 0;
QY 914 TGAATATTGTGAAATTAACCAAGAACTGGGAGAGAAATATCATAGAAAAAGGC 973
DB 34 TGGAAATAAGGTCGATGTTTCATAGCATCGTTCCAGTAGGACAAGATCCACATGAATATGA 93
QY 974 TATTGTTAAGGAAGTAATTGACAAATATACAGCTGTTGTGAAGATGATTGATTTCTGGAGA 1033
DB 94 GGTAAACCTAAAGATATTAAAGCATTTAACAGATGCTGACGTTGTTTATATATGTTT 153
QY 1034 CAAGCTGAACTTGACCAGACTCATTTAGAGACAGTAATTTCCAGACACAGAAAAAGAA 1093
DB 154 AAACCTAGAACTGGAAATGTTGTTTGAATAAGCACTTGACCAAGCAGGAAAAATCAAC 213
QY 1094 TCTAGTTTAAATGGAGGCTACAGAGGAATGAAGGTACCTAGATCCATCAATGAGAA 1153
DB 214 AAAAGATAAAAAATGTATAGCAGCATCAATAATGTTTAAACCAATATATATCTTAAATGGTGA 273
QY 1154 G 1154
DB 274 G 274

RESULT 4

US-10-793-626-2593/c
; Sequence 2593, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2593
; LENGTH: 2970
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-2593

Query Match 2.6%; Score 39; DB 1; Length 2970;
Best Local Similarity 46.2%; Pred. No. 1.8;
Matches 129; Conservative 0; Mismatches 150; Indels 0; Gaps 0;
QY 1249 TTGAAAAATTTGTTAACATATACATTAATAATCTTAAAGCATCAAAATGTTGTCGCAAGG 1308
DB 2444 TTGTTAAATCTCTATTATATATTTTAAACCTTAACTTCACTCAAGAAATCGTTTCATCACTA 2385
QY 1309 CATTATGAGACTCTACTGTGTTAGGTATATTTCTTTGTATATAAAACAACAGGTTTTTGA 1368
DB 2384 TTATATTATCTTTTAAACCGGAAGTAATTTTATGATATATAGATATTGGTACATTA 2325
QY 1369 AAATATTACTGTATAGTTGTTTCAGCTAACTTTGAGAGAAATTTAATTAATGTTCTCATGAG 1428
DB 2324 GTTTTAAGTACAATTTGTTTAAAAAACTTACCATAGTTTCATATAAATAAATTCATTCTTA 2265
QY 1429 GTATCAAACTATGTAATTTTCTCTCTGTTATTTTGTTCCTTTGTTTACTTTACTTGATGA 1488
DB 2264 TAATTATATCTGAAATAGATTTCTTGATACTTGTGTAACTTGTGTACCTTCATCACTTTTCCAAA 2205
QY 1489 GTTTATATCTTCAATTAAGAATGTTATTATATAAAAAAAA 1527
DB 2204 GATTGATGATTTCTTCTGTTGTTGTTTAAACAATAAA 2166

RESULT 5									
US-10-793-626-3796	; Sequence 3796, Application US/10793626								
	; Publication No. US20050255478A1								
	; GENERAL INFORMATION:								
	; APPLICANT: KIMMERLY, WILLIAM JOHN								
	; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS								
	; FILE REFERENCE: PU3480US								
	; CURRENT APPLICATION NUMBER: US/10/793,626								
	; CURRENT FILING DATE: 2004-03-04								
	; PRIOR APPLICATION NUMBER: 60/164,258								
	; PRIOR FILING DATE: 1999-11-09								
	; NUMBER OF SEQ ID NOS: 4472								
	; SOFTWARE: PatentIn Ver. 2.1								
	; SEQ ID NO 3796								
	; LENGTH: 3549								
	; TYPE: DNA								
	; ORGANISM: Artificial Sequence								
	; FEATURE:								
	; OTHER INFORMATION: Description of Artificial Sequence: synthetic								
	; OTHER INFORMATION: nucleic acid sequence								
	; US-10-793-626-3796								
	Query Match	2.6%	Score 39;	DB 1;	Length 3549;				
	Best Local Similarity	46.28;	Pred. No. 2;						
	Matches 129;	Conservative 0;	Mismatches 150;	Indels 0;	Gaps 0;				
QY	1249	TTTGAAAAATTTGTTAAACATACATATAAAATCTTTAAAGCATCAAAATGGTGTTCGCCAAGG	1308						
DB	2230	TTGTGTAATTTCTCTATTATATTATTATAAACTTAATTCACCTCAAGAATCGTTCATCACTA	2289						
QY	1309	CATTATCAGACTCTACTGTGTGGGTATATTCTTTTGTATAAACAAACAGGTTTTTGA	1368						
DB	2290	TTATATTATTCTTTTAAACCGAAAGTAAATTTTTATGCATTATAGATATTGGTACATTA	2349						
QY	1369	AAATATTACTGTATAGTTGTTTCAGCTAAACTTTTGAGAAAGAAATTTAAATATATGCTCATGAG	1428						
DB	2350	GTTTAAAGTACAATTTGTTTAAAACTTACCATAGTTTCATATAAATAAATTCATTCTA	2409						
QY	1429	GTATCAAACTATGTAATTTTGTCTCTGTTATTTTTTGTTCCTTTGTAATTTACTTTGATGA	1488						
DB	2410	TAATTATATTCTGAAATAGATTCTCTGATAACTTGTGTACCTTCATCATCTTTTTTCCAAA	2469						
QY	1489	GTTTATATCTTCATTAAGAATGTTATATAAAAAAAA	1527						
DB	2470	GATTGATGATCTCTTCTCGTTGTTGTTTAAACAAATAAA	2508						
RESULT 6									
US-10-793-626-1525/c	; Sequence 1525, Application US/10793626								
	; Publication No. US20050255478A1								
	; GENERAL INFORMATION:								
	; APPLICANT: KIMMERLY, WILLIAM JOHN								
	; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS								
	; FILE REFERENCE: PU3480US								
	; CURRENT APPLICATION NUMBER: US/10/793,626								
	; CURRENT FILING DATE: 2004-03-04								
	; PRIOR APPLICATION NUMBER: 60/164,258								
	; PRIOR FILING DATE: 1999-11-09								
	; NUMBER OF SEQ ID NOS: 4472								
	; SOFTWARE: PatentIn Ver. 2.1								
	; SEQ ID NO 1525								
	; LENGTH: 966								
	; TYPE: DNA								
	; ORGANISM: Artificial Sequence								
	; FEATURE:								
	; OTHER INFORMATION: Description of Artificial Sequence: synthetic								
	; OTHER INFORMATION: nucleic acid sequence								
	; US-10-793-626-1525								

```
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3753

Query Match          2.5%; Score 37.8; DB 1; Length 3782;
Best Local Similarity 53.8%; Pred. No. 4.1;
Matches 78; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 1368 AAAATATTACTGTATAGTTGTTTCAGCTAAACCTTCGAGAGAAATTTAATTATGCTCTCATGA 1427
Db 1585 ATAATCTTACTGTTTGTAGTACACCTAATGATCGATAAACTCTATCCAGTGTCTCTACA 1644

QY 1428 GGTATCAAACTATGTAATTTGCTGCTGTTATTTTGGTTTCCCTGTAATTTACTTTGATG 1487
Db 1645 GAATGTGTTTATCAAGCTTCTCTATTTGCTTTTCTTCAATTTATTAATTTGATTACA 1704

QY 1488 AGTTTATATCTTCATTAAGAATGT 1512
Db 1705 ACTTCAGTTAAGTTATCGATAATGT 1729

RESULT 9
US-10-793-626-3856
; Sequence 3856, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3856
; LENGTH: 3985
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3856

Query Match          2.5%; Score 37.8; DB 1; Length 3985;
Best Local Similarity 53.8%; Pred. No. 4.2;
Matches 78; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 1368 AAAATATTACTGTATAGTTGTTTCAGCTAAACCTTCGAGAGAAATTTAATTATGCTCTCATGA 1427
Db 3275 ATAATCTTACTGTTTGTAGTACACCTAATGATCGATAAACTCTATCCAGTGTCTCTACA 3334

QY 1428 GGTATCAAACTATGTAATTTGCTGCTGTTATTTTGGTTTCCCTGTAATTTACTTTGATG 1487
Db 3335 GAATGTGTTTATCAAGCTTCTCTATTTGCTTTTCTTCAATTTATTAATTTGATTACA 3394

QY 1488 AGTTTATATCTTCATTAAGAATGT 1512
Db 3395 ACTTCAGTTAAGTTATCGATAATGT 3419

RESULT 10
US-11-033-764-56
; Sequence 56, Application US/11033764
; Publication No. US2005024481A1
; GENERAL INFORMATION:
; APPLICANT: Rubin, Donald H.
; APPLICANT: Organ, Edward L.
; APPLICANT: DuBois, Raymond N.
; TITLE OF INVENTION: Mammalian Genes Involved in Viral

; TITLE OF INVENTION: Infection and Tumor Suppression
; FILE REFERENCE: 01123.0004
; CURRENT APPLICATION NUMBER: US/11/033.764
; CURRENT FILING DATE: 2005-01-12
; PRIOR APPLICATION NUMBER: US/10/877,807
; PRIOR FILING DATE: 2004-06-25
; PRIOR APPLICATION NUMBER: US/09/509,712
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: PCT/US98/21276
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/062,021
; PRIOR FILING DATE: 1997-10-10
; NUMBER OF SEQ ID NOS: 127
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 56
; LENGTH: 857
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1- 857
; OTHER INFORMATION: n = g, a, c or t(u)
US-11-033-764-56

Query Match          2.4%; Score 37.2; DB 9; Length 857;
Best Local Similarity 43.0%; Pred. No. 2.6;
Matches 96; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 435 AGAGAAGCGTTGTGCAAAAGTGGACGACACCAAAAGCGCTGTTATTTCTAGTACATAGAC 494
Db 44 ANAGANAGANAGAGNAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 103

QY 495 AGGACCCAGAAACTATCCCGCCGCAACTGGAACCTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 554
Db 104 GAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 163

QY 555 GATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 614
Db 164 NAAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 223

QY 615 AAGGAACAGGAGGTCCTCTACTTTTACGGAATTTAAGACAGAGAAA 657
Db 224 AANNAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 266

RESULT 11
US-10-793-626-4223
; Sequence 4223, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4223
; LENGTH: 3927
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4223

Query Match          2.4%; Score 37.2; DB 1; Length 3927;
Best Local Similarity 49.4%; Pred. No. 5.9;
Matches 154; Conservative 0; Mismatches 153; Indels 5; Gaps 2;

QY 1220 ATATGAGACATTTCTAAACTTCGCTGAGTTTGAAATTTGTTAAACAATACA---TTAAA 1276
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: November 26, 2005, 12:53:47 ; Search time 5857.21 Seconds
(without alignments)
10694.773 Million cell updates/sec

Title: US-09-555-529-2
Perfect score: 1102
Sequence: 1 atgggcagctcggtatttct.....aaaaaaaaaaaaaaaaaaaaa 1102

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba.*
2: gb_in.*
3: gb_env.*
4: gb_ov.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pr.*
9: gb_ro.*
10: gb_sts.*
11: gb_sy.*
12: gb_un.*
13: gb_vl.*
14: gb_htg.*
15: gb_pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1102	100.0	1102	6	AX003309	Sequence
2	799.2	72.5	1390	6	AX003331	Sequence
3	799.2	72.5	1414	9	NMKN17	X58472 Mouse KIN17
4	744.2	67.5	1386	9	BC028860	Mus muscu
5	704.6	63.9	1002	6	AX003310	Sequence
6	680	61.7	180882	14	AC114615	Mus muscu
7	680	61.7	182400	14	AC124010	Mus muscu
8	680	61.7	185378	9	AL772367	Mouse DNA
9	648.2	58.8	221547	14	AC119716	AL772367 Rattus no
10	611.8	55.5	2710	9	BC058169	Mus muscu
11	535	48.5	144118	14	AC141995	Rattus no
12	510.2	46.3	130262	14	AC166154	Mus muscu
13	510.2	46.3	140215	9	AC140205	Mus muscu
14	475	43.1	1528	6	AX003308	Sequence
15	475	43.1	1528	8	HSJ5273	AJ005273 Homo sapi
16	469	42.6	2069	8	BC017309	Homo sapi
17	333	30.2	333	6	AX003341	Sequence
18	329.2	29.9	1002	6	AX003311	Sequence

19	309	28.0	461	6	CQ684251	Sequence
20	302	27.4	1117	6	CQ722054	Sequence
21	261.6	23.7	333	6	AX003340	Sequence
22	258.8	23.5	1483	5	BC074606	Xenopus t
23	252	22.9	1501	5	BC079988	Xenopus l
24	198.6	18.0	175480	14	AC127074	Rattus no
25	198.6	18.0	247615	14	AC098061	Rattus no
26	190	17.2	1372	2	AK113340	Ciona int
27	176	16.0	177276	14	AC154247	Mus muscu
28	174	15.8	952	15	AY039912	Arabidops
29	174	15.8	1267	15	AY051011	Arabidops
30	174	15.8	1563	15	AF360132	Arabidops
31	174	15.8	87967	15	AC05223	Arabidops
32	173	15.7	1824	15	AK069396	Arabidops
33	173	15.7	101073	15	AC091302	Oryza sat
34	173	15.7	110000	15	AP083209	Continuation (205
35	173	15.7	152883	15	AC084319	Oryza sat
36	172	15.6	760	2	DME6528	Drosophill
37	168.8	15.3	1764	6	BD237772	Maize KIN
38	167.6	15.2	1780	6	BD237773	Maize KIN
39	167.6	15.2	1780	6	BD237774	Maize KIN
40	163	14.8	1241	6	CQ613281	Sequence
41	163	14.8	1467	2	AY084184	Drosophill
42	163	14.8	3313	6	CQ613280	Sequence
43	163	14.8	4394	6	CQ613259	Sequence
44	163	14.8	4582	6	CQ596774	Sequence
45	163	14.8	31807	14	AC017679	Drosophill

ALIGNMENTS

RESULT 1
AX003309
LOCUS AX003309
DEFINITION Sequence 2 from Patent WO9929845.
ACCESSION AX003309
VERSION AX003309.1 GI:9927126
KEYWORDS
SOURCE Mus sp.
ORGANISM Mus sp.

REFERENCE 1
AUTHORS Angulo-Mora, J.F. and Mauffrey, P.
TITLE Sequences coding for kin17 protein and their applications
JOURNAL Patent: WO 9929845-A 2 17-JUN-1999;
ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE (FR)

FEATURES
source Location/Qualifiers
1..1102
/organism="Mus sp."
/mol_type="unassigned DNA"
/db_xref="taxon:10095"

ORIGIN

Query Match 100.0%; Score 1102; DB 6; Length 1102;
Best Local Similarity 100.0%; Pred. No. 5e-200;
Matches 1102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGGCAAGTCGAGTTTCTGAGCCCAAGGCCATGCCCAATAGATTAAAGTCCAAAGGG 60
DB 1 ATGGGCAAGTCGAGTTTCTGAGCCCAAGGCCATGCCCAATAGATTAAAGTCCAAAGGG 60
QY 61 CTCAGAGCTTCGCTGGTACTGCCAGATGTGCGCAAGCAATGCGCGACGAGATGGC 120
DB 61 CTCAGAGCTTCGCTGGTACTGCCAGATGTGCGCAAGCAATGCGCGACGAGATGGC 120
QY 121 TTTAAGTGTCACTGTATGTCTGAATCTCATCAAGCAACTGTGTGCTTGCAGAAAC 180
DB 121 TTTAAGTGTCACTGTATGTCTGAATCTCATCAAGCAACTGTGTGCTTGCAGAAAC 180

QY 181 CCTCAGCAGTTTATGGAATTAATTTTCAGAGGAATTCGAAATGACTTTCTGGAACCTTCG 240
DB 181 CCTCAGCAGTTTATGGAATTAATTTTCAGAGGAATTCGAAATGACTTTCTGGAACCTTCG 240
QY 241 AGCGCAGCGTTTGGCAGTAAAGGGTCCCAACAACATTTGTCTACAAATGAATACATCAGC 300
DB 241 AGCGCAGCGTTTGGCAGTAAAGGGTCCCAACAACATTTGTCTACAAATGAATACATCAGC 300
QY 301 CACCGAGAGCACATCCACATGAAACGCTACCCAGTGGGAGACATGACGACCTTTACCAAG 360
DB 301 CACCGAGAGCACATCCACATGAAACGCTACCCAGTGGGAGACATGACGACCTTTACCAAG 360
QY 361 TGGCTGGGAGAGAGGGCTTGTAAAGTACCAAGTCCAGTGGGAGACATGACGACCTTTACCAAG 420
DB 361 TGGCTGGGAGAGAGGGCTTGTAAAGTACCAAGTCCAGTGGGAGACATGACGACCTTTACCAAG 420
QY 421 GCATCCGGGAAACGGAAAGAGTCTTCAACAGAGCTCCGCCAGCCTCGGAAGAGGAAG 480
DB 421 GCATCCGGGAAACGGAAAGAGTCTTCAACAGAGCTCCGCCAGCCTCGGAAGAGGAAG 480
QY 481 TCGGCCCTGGATGAGATCAATGAGTTCGAAAGGAAAGAAAGGACCGACCGACAGAC 540
DB 481 TCGGCCCTGGATGAGATCAATGAGTTCGAAAGGAAAGAAAGGACCGACCGACAGAC 540
QY 541 GCCTGTTACAGCGGGGATCGTTGTGAAATTAACGAAGAGCTTGGGAGAGAAATAT 600
DB 541 GCCTGTTACAGCGGGGATCGTTGTGAAATTAACGAAGAGCTTGGGAGAGAAATAT 600
QY 601 CACAAGAAAGGGTTCGTTAAGGAAGTGTGACAGGTTACACAGCTGTGTTAAGATG 660
DB 601 CACAAGAAAGGGTTCGTTAAGGAAGTGTGACAGGTTACACAGCTGTGTTAAGATG 660
QY 661 ACTGACTCTGGAGACAGGCTGAACTGGACCGACACTCATTTAGAGACAGTCAATTCGGCC 720
DB 661 ACTGACTCTGGAGACAGGCTGAACTGGACCGACACTCATTTAGAGACAGTCAATTCGGCC 720
QY 721 CCGGGGAAAGGGTTCAGTTTAAATGAGGCTACAGAGGAATGAAGGCACTCTCGAA 780
DB 721 CCGGGGAAAGGGTTCAGTTTAAATGAGGCTACAGAGGAATGAAGGCACTCTCGAA 780
QY 781 TCATCAATGAGAGGCTTTTTCAGCCAGTAGTCAATTTGAACTGGACCTTTGAAAGGA 840
DB 781 TCATCAATGAGAGGCTTTTTCAGCCAGTAGTCAATTTGAACTGGACCTTTGAAAGGA 840
QY 841 CCGAGAGTTGAGGTAATCAATATGAAGACATATCTAAACTTGTCTGAGTTTGAATAATTT 900
DB 841 CCGAGAGTTGAGGTAATCAATATGAAGACATATCTAAACTTGTCTGAGTTTGAATAATTT 900
QY 901 GATAACACACATTTGAACTGTGAAGCATCAAAATGGTGTAGCCAGGCACTGTGTAC 960
DB 901 GATAACACACATTTGAACTGTGAAGCATCAAAATGGTGTAGCCAGGCACTGTGTAC 960
QY 961 TCTACTGTGTAGGGATTTGTTTCTATTAATAAAAAAAAAAATCACTATTTAAATACT 1020
DB 961 TCTACTGTGTAGGGATTTGTTTCTATTAATAAAAAAAAAAATCACTATTTAAATACT 1020
QY 1021 AGTGAATAGTTGGGTAATTTATTAATAATAATCTATGTTTTTTTAAAGTGAATAAAAAA 1080
DB 1021 AGTGAATAGTTGGGTAATTTATTAATAATAATCTATGTTTTTTTAAAGTGAATAAAAAA 1080
QY 1081 AAAAAAAAAAAAAAAAAAAAAA 1102
DB 1081 AAAAAAAAAAAAAAAAAAAAAA 1102

RESULT 2
AX003331
LOCUS AX003331 1390 bp DNA linear PAT 24-AUG-2000
DEFINITION Sequence 24 from Patent WO929845.
ACCESSION AX003331
VERSION AX003331.1 GI:9927146
KEYWORDS Mus sp.

ORGANISM Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Angulo-Mora, J. F. and Mauffrey, P.
TITLE Sequences coding for kin17 protein and their applications
JOURNAL Patent: WO 929845-A 24 17-JUN-1999;
ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE (FR)
FEATURES
source 1..1390
/organism="Mus sp."
/mol_type="unassigned DNA"
/db_xref="taxon:10095"
ORIGIN
Query Match 72.5%; Score 799.2; DB 6; Length 1390;
Best Local Similarity 79.1%; Pred. No. 2.8e-142; Indels 288; Gaps 1;
Matches 1099; Conservative 0; Mismatches 3;
QY 1 ATGGCAAGTCGGATTTTCTGAGCCCAAGGCCATCGCAATAGAAATTAAGTCCAAAGGG 60
DB 1 ATGGCAAGTCGGATTTTCTGAGCCCAAGGCCATCGCAATAGAAATTAAGTCCAAAGGG 60
QY 61 CTCAGAAAGCTTCGCTGTACTGCCAGATGTGCCAAAAGCAATGCCGCGACGAGAAATGCG 120
DB 61 CTCAGAAAGCTTCGCTGTACTGCCAGATGTGCCAAAAGCAATGCCGCGACGAGAAATGCG 120
QY 121 TTTAAGTCTCACTGTATGCTGAAATCTCATCAAGACAACTGTTGCTGCTGCCTCAGAAAC 180
DB 121 TTTAAGTCTCACTGTATGCTGAAATCTCATCAAGACAACTGTTGCTGCTGCCTCAGAAAC 180
QY 181 CCTCAGCAGTTTATGGAATTAATTTTTCAGAGGAATTCGAAATGACTTTCTGGAACCTTCG 240
DB 181 CCTCAGCAGTTTATGGAATTAATTTTTCAGAGGAATTCGAAATGACTTTCTGGAACCTTCG 240
QY 241 AGCGCAGCGTTTGGCAGTAAAGGGTCCCAACAACATTTGTCTACAAATGAATACATCAGC 300
DB 241 AGCGCAGCGTTTGGCAGTAAAGGGTCCCAACAACATTTGTCTACAAATGAATACATCAGC 300
QY 301 CACCGAGAGCACATCCACATGAAACGCTACCCAGTGGGAGACACTGACGACCTTTACCAAG 360
DB 301 CACCGAGAGCACATCCACATGAAACGCTACCCAGTGGGAGACACTGACGACCTTTACCAAG 360
QY 361 TGGCTGGGAGAGAGGGCTTGTGTAATAAGTGGATGAGACACCGAAAGGCTGTGTACATTGAG 420
DB 361 TGGCTGGGAGAGAGGGCTTGTGTAATAAGTGGATGAGACACCGAAAGGCTGTGTACATTGAG 420
QY 388 ----- 387
DB 421 TACATAGACAGAGACCCAGAAACCATCCGTCGGCAACTGGAATTAGAAAAAAGAAAGAG 480
QY 388 ----- 387
DB 481 CAAGATCTGGACGATGAAGAAAAAACTGCCAAGTTCAATTGAGGACGAGGTGAGAAAGAGC 540
QY 388 ----- 387
DB 541 CTGGAAGGGAAGAGCAGGAGACACCTGTTTACAGAACTTAGCCGAGAAAAATGAGGAA 600
QY 388 ----- 387
DB 601 GAAAAAGTTACGTTCAATCTGAATAAAGAGCGGGTGGCTCAGCGGGAGCTACAACTCC 660
QY 388 -----GGTACCAGTGCCTGAAGCTGCTGGGAGCGCGACATCCCGGAAA 432
DB 661 AAGTCAAGCTCTTTGGGACCAAGTGCATGAAGCTGCTGGGAGCGCGACATCCCGGAAA 720
QY 433 CGGAAAGAGTCTTCAACAGAGCTCCGCCAGCGCTCGAAAGAAAGAAAGTTCGCGCCTCGAT 492
DB 721 CGGAAAGAGTCTTCAACAGAGCTCCGCCAGCGCTCGGAAAGAAAGAAAGTTCGCGCCTCGAT 780

Db 205 CCTCAGCAGTTATGGATTATTTTTTCAGAGGAATTCGGAATGACTTTCTGGAACTTCG 264
Qy 241 AGCGCAGCTTTGGCACTAAAAAGGTCCTCAACAACATTTGTCTACAAATGAATACATCAGC 300
Db 265 AGCGCAGCTTTGGCACTAAAAAGGTCCTCAACAACATTTGTCTACAAATGAATACATCAGC 324
Qy 301 CACCGAGACCATCCACATGAACGCTACCCAGTGGGAGACACTGACCCAGCTTTACCAAG 360
Db 325 CACCGAGACCATCCACATGAACGCTACCCAGTGGGAGACACTGACCCAGCTTTACCAAG 384
Qy 361 TGGCTGGGCGAGAGGGCTTTGTAA----- 387
Db 385 TGGCTGGGCGAGAGGGCTTTGTAAAGTGGATGACACACCGAAGGCTGTACATTGAG 444
Qy 388 ----- 387
Db 445 TACATAGACAGAGCCAGAAACCATCCCTCGCAACTGGAATTAGAAAAAAGGAAG 504
Qy 388 ----- 387
Db 505 CAAGATCTGGACGATGAAGAAAAAACTGCCAAGTTCAATTAGGAGCAGGTGAGAAAGGC 564
Qy 388 ----- 387
Db 565 CTGGAGGAAAGAGCAGGAGACACCTGTTTTACAGAATTAGCGGAGAAATGAGGAA 624
Qy 388 ----- 387
Db 625 GAAAAAGTTACGTTCAATCTGAATAAAGGAGCGGTGGCTCAGCGGAGCTACAACATCC 684
Qy 388 -----GTTACCAAGTCACTGAAGAGCTGCTGGGAGCGCAGCATCCGGGAAA 744
Db 685 AAGTCAAGCTCTTTGGGACCAAGTCACTGAAGAGCTGCTGGGAGCGCAGCATCCGGGAAA 744
Qy 433 CGGAAAGAGCTTTCAAGAGCTCCGCCAGCTCGCAAGAGAGTGGCCCTCGGAT 492
Db 745 CGGAAAGAGCTTTCAAGAGCTCCGCCAGCTCGCAAGAGAGTGGCCCTCGGAT 804
Qy 493 GAGATCATGGAGCTCGAAGAGGAAAGAAAGGACCGCAGCAGACGCCCTGGTTACAG 552
Db 805 GAGATCATGGAGCTCGAAGAGGAAAGAAAGGACCGCAGCAGACGCCCTGGTTACAG 864
Qy 553 CCGGGATCGTTGTGAAATTTATACGAAGAGAGCTTTGGGAGAAATATACAAAGAGAA 612
Db 865 CCGGGATCGTTGTGAAATTTATACGAAGAGAGCTTTGGGAGAAATATACAAAGAGAA 924
Qy 613 GGGGTCGTTAAGAGAGTGTACAGGTACACAGCTGTGTAAAGATGACTCTCTGGA 672
Db 925 GGGGTCGTTAAGAGAGTGTACAGGTACACAGCTGTGTAAAGATGACTCTCTGGA 984
Qy 673 GACAGGCTGAAACTGGACAGACTCATTATGAGACAGTCAATTCGCGCCCGGGGAAAAGG 732
Db 985 GACAGGCTGAAACTGGACAGACTCATTATGAGACAGTCAATTCGCGCCCGGGGAAAAGG 1044
Qy 733 GTTCTAGTTTTAAATGGAGGCTACAGAGAAATGAAGGCACTCTCGAATCCATCAATGAG 792
Db 1045 GTTCTAGTTTTAAATGGAGGCTACAGAGAAATGAAGGCACTCTCGAATCCATCAATGAG 1104
Qy 793 AAGGCTTTTTCAGCCACGATGATCAATTGAAACTGGACCTTCAAGAGGAGCGCAGAGTTGAA 852
Db 1105 AAGGCTTTTTCAGCCACGATGATCAATTGAAACTGGACCTTCAAGAGGAGCGCAGAGTTGAA 1164
Qy 853 GGTATTTCAATATGAAGACATATCTAACTGCTTGTGATTTGAAATTTGATTAACAACACA 912
Db 1165 GGTATTTCAATATGAAGACATATCTAACTGCTTGTGATTTGAAATTTGATTAACAACACA 1224
Qy 913 TTGAACCTGTGAAGCATCAAAATGGTGTAGCCAGGCACTGTGTAACTCTACTGTGTTA 972
Db 1225 TTGAACCTGTGAAGCATCAAAATGGTGTAGCCAGGCACTGTGTAACTCTACTGTGTTA 1284
Qy 973 GGGGATTTCTTTGATTTAAAAAATAAATAATCTATTTAAATACTAGTGAATGTTG 1032
Db 1285 GGGGATTTCTTTGATTTAAAAAATAAATAATCTATTTAAATACTAGTGAATGTTG 1344

Qy 1033 GGTAAATTTATAATAAAATCTATGTTTTTTTAAAGTAAATAAAAAAAAAAAAAA 1092
Db 1345 GGTAAATTTATAATAAAATCTATGTTTTTTTAAAGTAAATAAAAAAAAAAAAAA 1404
Qy 1093 AAAAAAAAAA 1102
Db 1405 AAAAAAAAAA 1414

RESULT 4
BC028860
LOCUS
DEFINITION
Mus musculus antigenic determinant of rec-A protein, mRNA (CDNA
clone MGC:18413 IMAGE:3673830), complete cds.
ACCESSION
BC028860
VERSION
BC028860.1 GI:22135637
KEYWORDS
MGC.
SOURCE
Mus musculus (house mouse)
ORGANISM

REFERENCE
AUTHORS
1 (bases 1 to 1386)
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Dietzen, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Udén, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.E.,
Scherer, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 1386)
Strausberg, R.
Direct Submission
Submitted (01-MAY-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louleeged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.

TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

REMARK
COMMENT
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louleeged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LINL at: <http://image.llnl.gov>
Series: IRAK Plate: 23 Row: h Column: 8
This clone was selected for full length sequencing because it
passed the following selection criteria: Genomescan gene

FEATURES		prediction.	
source		Location/Qualifiers	
		1..1386	
		/organism="Mus musculus"	
		/mol_type="mRNA"	
		/strain="FVB/N"	
		/db_xref="taxon:10090"	
		/clone="MGC:18413 IMAGE:3673830"	
		/tissue_type="Mammary tumor; C3(1)-Tag model. Infiltrating ductal carcinoma. 5 month old virgin mouse."	
		/clone_lib="NCI CGAP_Mam6"	
		/lab_host="DH10B"	
		/note="Vector: pCMV-SPORT6"	
gene		1..1386	
		/gene="Kin"	
		/note="synonym: Kin17"	
		/db_xref="GeneID:16588"	
		/db_xref="MGI:96676"	
		20..1195	
		/gene="Kin"	
		/codon_start=1	
		/product="antigenic determinant of rec-A protein"	
		/protein_id="AAH28860.1"	
		/db_xref="GI:22135638"	
		/db_xref="GeneID:16588"	
		/db_xref="MGI:96676"	
		/translation="MGKSDPLSPKAIANRIKSKLOLRWYCMQCKQCDENGFKCH CMSEHQRLLLASENPQPMDFSEFRNDFLELRFRPGTKRVNNIVYNIYISHR EHTHMAATWETLTDTKWLREGCLKVDETPKGWIQVIDEKPETIRRLQLEKKKK QDLDDDEKTAFT EEOVRGLEGEQETPVFTELSRENBEKEKVTFLNKGAGGSAGT TSKSSISGPSALXLLGSAAGKRKSSQSAOPAKKKSALDSEIMLEBEKKRTATD AWLQPGIVKILTKLGEKXHKYKGVKVIIDRYTAVKWTDSGDLKLDOTHLETVI PAPGKRVLVNGGIRNGEGTLESINEKAFSATIVIEGTPLKRRVREGIQYEDISKLA"	
CDS			
		67.5%; Score 744.2; DB 9; Length 1386;	
		Query Match 77.6%; Pred. No. 8.9e-132;	
		Best Local Similarity 0; Mismatches 13; Indels 294; Gaps 2;	
		Matches 1066; Conservative	
Qy	1	ATGGGCAAGTCGGAATTTCTGAGCCCAAGCCCATGCCAATAGAATTAAGTCAAAGGG	60
Db	20	ATGGGCAAGTCGGAATTTCTGAGCCCAAGCCCATGCCAATAGAATTAAGTCAAAGGG	79
Qy	61	CTCCAGAAGCTTCGCTGGTACTGCGAGATGCGCAAGCAATGCCCGCAGCAATGGC	120
Db	80	CTCCAGAAGCTTCGCTGGTACTGCGAGATGCGCAAGCAATGCCCGCAGCAATGGC	139
Qy	121	TTTAAAGTGTCATGTCATGTCATCAAGACAACTGTTGCTGGCTTCAGAAAC	180
Db	140	TTTAAAGTGTCATGTCATGTCATCAAGACAACTGTTGCTGGCTTCAGAAAC	199
Qy	181	CCTCAGCAGTTTATGGAATATTTTTCAGAGGAATTCGGAATAGCTTTCTGGAATCTTG	240
Db	200	CCTCAGCAGTTTATGGAATATTTTTCAGAGGAATTCGGAATAGCTTTCTGGAATCTTG	259
Qy	241	AGGCGACGTTTGCGCATAAAGGGTCCACAACTTGTCTACAAATGAATACATCAGC	300
Db	260	AGGCGACGTTTGCGCATAAAGGGTCCACAACTTGTCTACAAATGAATACATCAGC	319
Qy	301	CACCGAGGACATCCACATGAACGCTACCAGTGGGAGACACTGACCGACTTTTACCAG	360
Db	320	CACCGAGGACATCCACATGAACGCTACTCAGTGGGAGACACTGACCGACTTTTACCAG	379
Qy	361	TGGCTGGCGCAGAGGGCTTGTGTAATAA	387
Db	380	TGGCTGGCGCAGAGGGCTTGTGTAATAA	439
Qy	388		387
Db	440	TACATAGACAGACCCAGAAACCATCCGTCGGCAACTGGAAATAGAAAAAGAAAG	499
Qy	388		387
Db	500	CAAGATCTGGACGATGAAGAAAAAACTGCCAAGTTTCATTGAGGACAGCGTGAGAGAGGC	559
Qy	388		387
Db	560	CTGGGAAGGAAAGACGAGGAGACACCTGTTTTTACAGAACTAAGCCGAGAAATGAGGAA	619
Qy	388		387
Db	620	GAAGAAGTTACGTTCAATCTGAATAAAGGAGCGGGTGGCTCAGCGGGAGCTACCAATCC	679
Qy	388		387
Db	680	AAAGTCAAGCTCTTTGGGACCAAGTGCACCTGAAGCTCTGGGAGCCGAGCATCCGGGAAA	739
Qy	433	CGGAAAGAGTCTTTCACAGAGCTCCGCCAGCTCGGAGGAAAGAGTCCGCCCTGGAT	492
Db	740	CGGAAAGAGTCTTTCACAGAGCTCCGCCAGCTCGGAGGAAAGAGTCCGCCCTGGAT	799
Qy	493	GAGATCATCGAGCTCGAAGAGGAAAAAGAACCGGACCGGACAGACGCTCGTTACAG	552
Db	800	GAGATCATCGAGCTCGAAGAGGAAAAAGAACCGGACCGGACAGACGCTCGTTACAG	859
Qy	553	CCGGGATCGTTTGTGAAATTTATACGAAGAAGCTTTGGGAGGAAATATACAGAAGAAA	612
Db	860	CCGGGATCGTTTGTGAAATTTATACGAAGAAGCTTTGGGAGGAAATATACAGAAGAAA	919
Qy	613	GGGTCGTTAAAGAGTGTGACAGGTACACAGCTGTGTTAAAGATGACTGACTCTGGA	672
Db	920	GGGTCGTTAAAGAGTGTGACAGGTACACAGCTGTGTTAAAGATGACTGACTCTGGA	979
Qy	673	GACAGCTGAAACTGGACCCAGACTCATTTAGAGACAGTCAATCCGCCCCCGGGGAAAAGG	732
Db	980	GACAGCTGAAACTGGACCCAGACTCATTTAGAGACAGTCAATCCGCCCCCGGGGAAAAGG	1039
Qy	733	GTTCATGTTTAAATGGAGCTACAGAGAAATGAAGGCACTCTCGAATCCATCAATGAG	792
Db	1040	GTTCATGTTTAAATGGAGCTACAGAGAAATGAAGGCACTCTCGAATCCATCAATGAG	1099
Qy	793	AGGCTTTTTCAGCCACAGTAGTCATTGAAACTGGACCTTTGAAAGGACGACAGAGTTGAA	852
Db	1100	AGGCTTTTTCAGCCACAGTAGTCATTGAAACTGGACCTTTGAAAGGACGACAGAGTTGAA	1159
Qy	853	GGTATTCAATATGAAGACATATCTAAACTTGTCTGAGTTTGAATAATTTGATAACAACACA	912
Db	1160	GGTATTCAATATGAAGACATATCTAAACTTGTCTGAGTTTGAATAATTTGATAACAACACA	1219
Qy	913	TTGAAACTGTGAAGCATCAAAATGGTGTAGCCAAAGGCACTGTGTAACTCTACTGTGTTA	972
Db	1220	TTGAAACTGTGAAGCATCAAAATGGTGTAGCCAAAGGCACTGTGTAACTCTACTGTGTTA	1279
Qy	973	GGGATTTGTTTGTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	1032
Db	1280	GGGATTTGTTTGTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	1333
Qy	1033	GGTAAATTTATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	1085
Db	1334	GGTAAATTTATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	1386
RESULT 5			
AX003310		1002 bp	DNA
LOCUS		Sequence 3 from Patent WO9929845.	linear
DEFINITION		AX003310	
ACCESSION		AX003310.1	
VERSION		GI:9927127	
KEYWORDS		Homo sapiens (human)	
SOURCE		Homo sapiens	
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE		1	
AUTHORS		Angulo-Mora, J.F. and Mauffrey, P.	

JOURNAL Submitted (25-AUG-2002) Genome Sequencing Center, 4414 Forest Park Parkway, St. Louis, MO 63108, USA

COMMENT On Aug 25, 2002 this sequence version replaced qi:21327639.

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc/index.shtml>
 Contact: submissions@watson.wustl.edu
 ----- Project Information -----
 Center project name: M BA0119N04

Sequencing vector: pMI3; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.950319
Consensus quality: 183528 bases at least Q40
Consensus quality: 184210 bases at least Q30
Consensus quality: 184756 bases at least Q20
Insert size: 197000; agarose-fp
Insert size: 196816; sum-of-contigs
Quality coverage: 9.51 in Q20 bases; agarose-fp
Quality coverage: 9.61 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of 'N', but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be changed.

*	1	11868:	contig of 11868 bp in length
*	11869	11968:	gap of unknown length
*	11969	28906:	contig of 16938 bp in length
*	28907	29006:	gap of unknown length
*	29007	72773:	contig of 43767 bp in length
*	72774	72873:	gap of unknown length
*	72874	124926:	contig of 52053 bp in length
*	124927	125026:	gap of unknown length
*	15027	181902:	contig of 56876 bp in length
*	181903	182002:	gap of unknown length
*	182003	182400:	contig of 398 bp in length.

FEATURES	source	Location/Qualifiers
		1..182400
		/organism="Mus musculus"
		/mol_type="genomic DNA"
		/db_xref="taxon:10090"
		/chromosome="UNK"
		/clone="RP23-119N4"
misc_feature		1..11868
		/note="assembly_name:Contig85"
gap		11869..11968
		/estimated_length=unknown
misc_feature		11969..28906
		/note="assembly_name:Contig87"
gap		28907..29006
		/estimated_length=unknown
misc_feature		29007..72773
		/note="assembly_name:Contig88"
gap		72774..72873
		/estimated_length=unknown
misc_feature		72874..124926
		/note="assembly_name:Contig89"
gap		124927..125026
		/estimated_length=unknown
misc_feature		125027..181902
		/note="assembly_name:Contig90"
gap		181903..182002
		/estimated_length=unknown
misc feature		182003..182400

ORIGIN	/note="assembly_name:Contig16"				
Query Match	61.7%;	Score 680;	DB 14;	Length 182400;	
Best Local Similarity	99.3%;	Pred. No. 8.7e-120;			
Matches 683;	Conservative 0;	Mismatches 5;	Indels 0;	Gaps 0;	
Qy	388	GGTACCAGTGCAC	TGAAGCTGCTGGGAGCGCAGCATCCGGAAACGGAAGAGTCTTCA	447	
Db	168848	GGACCAAGTGCAC	TGAAGCTGCTGGGAGCGCAGCATCCGGAAACGGAAGAGTCTTCA	168907	
Qy	448	CAGAGCTCCGCCAGCCTCGGAAGAGAAAGTCGGCCCTCGATGAGATCATGGAGCTC	507		
Db	168908	CAGAGCTCCGCCAGCCTCGGAAGAGAAAGTCGGCCCTCGATGAGATCATGGAGCTC	168967		
Qy	508	GAAGAGGAAAGAAAGGACCGCAGCAGACGCGCTGGTTTACAGCCGGGATCGTTCTG	567		
Db	168968	GAAGAGGAAAGAAAGGACCGCAGCAGACGCGCTGGTTTACAGCCGGGATCGTTCTG	169027		
Qy	568	AAATATTATAACGAAGAGCTTTGGGGAGAAATATCACAAAGAAAGAGGGTCGTTTAAGNA	627		
Db	169028	AAATATTATAACGAAGAGCTTTGGGGAGAAATATCACAAAGAAAGAGGGTCGTTTAAGNA	169087		
Qy	628	GTGATTGACAGGTACACAGCTGTGGTAAAGATGACTGCTCTGGAGACAGGCTGAAACTG	687		
Db	169088	GTGATTGACAGGTACACAGCTGTGGTAAAGATGACTGCTCTGGAGACAGGCTGAAACTG	169147		
Qy	688	GACCAGACTCATTTAGAGACAGTCATCCCGCCCCGGGGAAAGGGTCTAGTTTAAAT	747		
Db	169148	GACCAGACTCATTTAGAGACAGTCATCCCGCCCCGGGGAAAGGGTCTAGTTTAAAT	169207		
Qy	748	GGAGGCTACAGAGGAATGAAGCAGCTCTCGAATCCATCAATGAGAAGGCTTTTTTCAGCC	807		
Db	169208	GGAGGCTACAGAGGAATGAAGCAGCTCTCGAATCCATCAATGAGAAGGCTTTTTTCAGCC	169267		
Qy	808	ACGATAGTCAATTGAAAACGTGCAGCCTTTGAAAGGCGCAGAGTTGAAGGTATTCAATATGAA	867		
Db	169268	ACGATAGTCAATTGAAAACGTGCAGCCTTTGAAAGGCGCAGAGTTGAAGGTATTCAATATGAA	169327		
Qy	868	GACATATCTAAAACCTTGCTTGAGTTTGAAAATTTTGATTAACAACAACATTTGAAACCTGTGAAGC	927		
Db	169328	GACATATCTAAAACCTTGCTTGAGTTTGAAAATTTTGATTAACAACAACATTTGAAACCTGTGAAGC	169387		
Qy	928	ATCAAAATTGGTGTAGCCAGGCACTGTGTAACTCTACTCTGTTAGGGGATTCGTTTGT	987		
Db	169388	ATCAAAATTGGTGTAGCCAGGCACTGTGTAACTCTACTCTGTTAGGGGATTCGTTTGT	169447		
Qy	988	ATTAAAAAATAAATAATCATCTATTTAAATACTAGTGAATAGTTGGGTAAATTTTATAATA	1047		
Db	169448	ATTAAAAAATAAATAATCATCTATTTAAATACTAGTGAATAGTTGGGTAAATTTTATAATA	169507		
Qy	1048	AAATCTATGTTTTTTTTTAAAGTGTAAAAA	1075		
Db	169508	AAATCTATGTTTTTTTTTAAAGTGTAAAAA	169515		

RESULT	8
AL772367/c	
LOCUS	AL772367 Mouse DNA sequence from clone R223-119N4 on chromosome 2, complete sequence.
DEFINITION	Mouse DNA sequence from clone R223-119N4 on chromosome 2, complete sequence.
ACCESSION	AL772367
VERSION	AL772367.7 GI:22759512
KEYWORDS	HTG.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus. 1 (bases 1 to 183378) Wood,J.
AUTHORS	Direct Submission
TITLE	Submitted (30-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
JOURNAL	

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Sep 9, 2002 this sequence version replaced gi:22625488.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Web site: <http://www.sanger.ac.uk>
 Contact: humquery@sanger.ac.uk

 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em., EMBL; Sw.,
 SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP
 database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-119N4 is
 from the RPCI-23 Mouse PAC library
 constructed by the group of Pieter de Jong.
 For further details see <http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACe3.6. Qualifiers
 1. 185378
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="2"
 /clone="RP23-119N4"
 /clone_lib="RPCI-23"

FEATURES

source

ORIGIN

Query Match 61.7% Score 680; DB 9; Length 185378;
 Best Local Similarity 99.3%; Pred. No. 8.7e-120;
 Matches 683; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 388 GGTACCAAGTGCATGAAGCTGCTGGGAGCGCAGCATCCGGGAAACGGAAGACTTTCA 447
 DB 139979 GGACCAAGTGCATGAAGCTGCTGGGAGCGCAGCATCCGGGAAACGGAAGACTTTCA 139920
 QY 448 CAGAGCTCCGCCAGCTCGGAAGAAGAAAGTCCGCCCTCGATGATGATCATGGAGCTC 507
 DB 139919 CAGAGCTCCGCCAGCTCGGAAGAAGAAAGTCCGCCCTCGATGATGATCATGGAGCTC 139860
 QY 508 GAAGAGGAAGAAAGGACCGCAGCAGCAGCCCTGTTACAGCCGGGATCGTTGTG 567
 DB 139859 GAAGAGGAAGAAAGGACCGCAGCAGCAGCCCTGTTACAGCCGGGATCGTTGTG 139800
 QY 568 AAAATTATTAACGAAGAAGCTTGGGAGAAATATCACAAGAAAGAAAGGGTCTGTTAAGGAA 627
 DB 139799 AAAATTATTAACGAAGAGCTTGGGAGAAATATCACAAGAAAGAAAGGGTCTGTTAAGGAA 139740
 QY 628 GTGATTGACAGGTCACAGCTGTGGTAAAGATGATGATCTCTGGAGACAGCTGAAGCTG 687
 DB 139739 GTGATTGACAGGTCACAGCTGTGGTAAAGATGATGATCTCTGGAGACAGCTGAAGCTG 139680
 QY 688 CACAGACTCATTTAGACAGACTCATTCGCGCCCGCGGAAAGGGTCTAGTTTTAAAT 747
 DB 139679 CACAGACTCATTTAGACAGACTCATTCGCGCCCGCGGAAAGGGTCTAGTTTTAAAT 139620
 QY 748 GGAGGCTACAGAGGAATGAAGGACCTCTCGAATCCCATTAATGAGAAGGCTTTTTCAGCC 807
 DB 139619 GGAGGCTACAGAGGAATGAAGGACCTCTCGAATCCCATTAATGAGAAGGCTTTTTCAGCC 139560

QY 808 ACATAGTACATTGAAACCTGGACCTTTGAAAGGACCCAGAGTTGAAGGTATTCAATATGAA 867
 DB 139559 ACATAGTACATTGAAACCTGGACCTTTGAAAGGACCCAGAGTTGAAGGTATTCAATATGAA 139500
 QY 868 GACATATCTAAACTTGGCTTGGCTTGGAAATTTGATACACACATTTGAAACCTGGAAGC 927
 DB 139499 GACATATCTAAACTTGGCTTGGCTTGGAAATTTGATACACACATTTGAAACCTGGAAGC 139440
 QY 928 ATCAAAATTTGGCTTAGCCAAAGGCACCTGTGTAACCTCTACTGTGTAGGGGATTTGTTTGT 987
 DB 139439 ATCAAAATTTGGCTTAGCCAAAGGCACCTGTGTAACCTCTACTGTGTAGGGGATTTGTTTGT 139380
 QY 988 ATTAAAAAAAATAATCATCTATTAAATATCTAGTGAATAGTTGGGTAAAATTTATAATA 1047
 DB 139379 ATTAAAAAAAATAATCATCTATTAAATATCTAGTGAATAGTTGGGTAAAATTTATAATA 139320
 QY 1048 AAATCATGTTTTTTTTTAAGTGTAAAAA 1075
 DB 139319 AAATCATGTTTTTTTTTAAGTGTAAAAA 139292
 RESULT 9
 AC119716 221547 bp DNA linear HTG 19-NOV-2002
 LOCUS Rattus norvegicus clone CH230-274116, *** SEQUENCING IN PROGRESS
 DEFINITION *** 17 unordered pieces.
 AC119716
 AC119716.6 GI:25092366
 HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
 KEYWORDS Rattus norvegicus (Norway rat)
 SOURCE Rattus norvegicus
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Murioidea; Murinae; Rattus.
 1 (bases 1 to 221547)
 Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gabregiorge, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
 Hollins, B., Howells, S., Huliy, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Loreshuhwa, L., Loulisedge, H., Lozano, R. J., Lu, X., Ma, J.,
 Mahendhara, M., Mahindratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naif, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
 Nwokeleneh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L. L.,
 Puzao, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
 Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,

Shetty J., Shvartbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Stealmie, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uemani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 221547)
Worley, K.C.

Direct Submission
Submitted (01-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 221547)

Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:23611286.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GWLC
Center clone name: CH230-274116
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 208013 bases at least Q40
Consensus quality: 211685 bases at least Q30
Consensus quality: 213907 bases at least Q20
Estimated insert size: 199182; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 17 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 198201: contig of 198201 bp in length
* 198202
* 198302
* 199749: contig of 1448 bp in length
* 199750
* 199849: gap of unknown length
* 199850
* 201115: contig of 1266 bp in length
* 201116
* 201215: gap of unknown length
* 201216
* 202338: contig of 1123 bp in length
* 202339
* 202438: gap of unknown length
* 202439
* 203500: contig of 1062 bp in length

* 203501 203600: gap of unknown length
* 203601 204905: contig of 1305 bp in length
* 204906 205005: gap of unknown length
* 205006 206192: contig of 1187 bp in length
* 206193 206292: gap of unknown length
* 206293 207385: contig of 1093 bp in length
* 207386 207485: gap of unknown length
* 207486 208519: contig of 1034 bp in length
* 208520 208619: gap of unknown length
* 208620 209758: contig of 1139 bp in length
* 209759 209858: gap of unknown length
* 209859 211441: contig of 1583 bp in length
* 211442 211541: gap of unknown length
* 211542 212876: contig of 1335 bp in length
* 212877 212976: gap of unknown length
* 212977 214256: contig of 1280 bp in length
* 214257 214356: gap of unknown length
* 214357 217003: contig of 2647 bp in length
* 217004 217103: gap of unknown length
* 217104 218459: contig of 1356 bp in length
* 218460 218559: gap of unknown length
* 218560 219788: contig of 1229 bp in length
* 219789 219888: gap of unknown length
* 219889 221547: contig of 1659 bp in length.

FEATURES
Location/Qualifiers
source
1..221547
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-274116"
misc_feature
1..1272
/note="wgs_end_extension
clone_end:Sp6"
misc_feature
complement(4416..5263)
/note="clone_boundary
clone_end:Sp6"
misc_feature
site:
end_sequence:BZ266352"
192445..193553
/note="wgs_contig"
195743..196592
/note="clone_boundary
clone_end:T7"
site:
end_sequence:BZ266349"
198202..198301
/estimated_length=unknown
199750..199849
/estimated_length=unknown
201116..201215
/estimated_length=unknown
202339..202438
/estimated_length=unknown
203501..203600
/estimated_length=unknown
204906..205005
/estimated_length=unknown
206193..206292
/estimated_length=unknown
207386..207485
/estimated_length=unknown
208520..208619
/estimated_length=unknown
209759..209858
/estimated_length=unknown
211442..211541
/estimated_length=unknown
212877..212976
/estimated_length=unknown
214257..214356
/estimated_length=unknown
217004..217103
/estimated_length=unknown

gap	218460. .218559	58.8%; Score 648.2; DB 14; Length 221547;	
	/estimated_length=unknown	70.6%; Pred. No. 9.9e-114;	
gap	219789. .219888	0; Mismatches 108; Indels 288; Gaps 1;	
	/estimated_length=unknown		
Query Match			
Best Local Similarity			
Matches 951; Conservative			
QY	1	ATGGGCAAGTCGGATTTCTGAGGCCCAAGGCCATCGCAATAGAAATTAAGTCCAAAGGG 60	
DB	213012	ATGGGCAAGTCGGATTTCTGAGCCCCCAAGGCCATCGCAATAGAAATTAAGTCCAAAGGG 213071	
QY	61	CTCCAGAAGCTTCGCTGGTACTGCCAGATGTGCCAAAAGCAATGCCGCGACGAGAATGGC 120	
DB	213072	CTCCAGAAGCTTCGCTGGTACTGCCAGATGTGCCAAAAGCAATGCCGCGACGAGAATGGC 213131	
QY	121	TTTAAAGTGTCATGTATGTCTGAATCTCATCAAGACAACTGTTCGCTGGCTTCAGAAAAC 180	
DB	213132	TTTAAAGTGTCATGTATGTCTGAATCTCATCAAGACAACTGTTCGCTGGCTTCAGAAAAC 213191	
QY	181	CCTCAGCAGTTTATGGATTTATTTTTCAGAGGAATTCGGAATAGCTTTCTGGAATCTCTG 240	
DB	213192	CCTCAGCAGTTTATGGATTTATTTTTCAGAGGAATTCGGAATAGCTTTCTGGAATCTCTG 213251	
QY	241	AGCGCAGCTTTTGGCACTAAAAGGGTCCACAACAACATTTGTCTCAATGAATACATCAGC 300	
DB	213252	AGCGCAGCTTTTGGCACTAAAAGGGTCCACAACAACATTTGTCTCAATGAATACATCAGC 213311	
QY	301	CACCGAGACATCCACATGAACGCTACCCAGTGGGAGACACTGACCGACTTTACCAAG 360	
DB	213312	CACCGAGACATCCACATGAACGCTACCCAGTGGGAGACACTGACCGACTTTACCAAG 213371	
QY	361	TGCTGGCGAGAGGGCTTGTGTAA----- 387	
DB	213372	TGCTGGCGAGAGGGCTTGTGTAAAGTGSATGAGACACCGAAAAGCGTGTACATTGAG 213431	
QY	388	----- 387	
DB	213432	TACATGACAGACCCAGAAACCATCCGTCGGCACTCGAATTAGAAAAAAGAAAG 213491	
QY	388	----- 387	
DB	213492	CAAGATCTGGACCATGAAGAAAAAACTGCCAAGTTTCATTGAGGAGCAGGTGAGAGAGGC 213551	
QY	388	----- 387	
DB	213552	CTGGAAGGGAAAGAGCAGGAGACACCTGTTTTTACAGAACTAAGCCGAGAAAATCAGGAA 213611	
QY	388	----- 387	
DB	213612	GAAAAAATTACGTTCTGAATCTGAATAAAGGAGCGGTGGCTCAGCGGAGCTCAACATCC 213671	
QY	388	-----GTTACCGTCACTGAAGCTCTCGGGAGCGCAGCATCCGGGAAA 432	
DB	213672	AAGTCAAGCTCTTTGGGACCAAGTGCACTGAAGCTCTCGGGAGCGCAGCATCCGGGAAA 213731	
QY	433	CGGAAAGAGTCTTCACAGAGCTCCGCCAGCTTCGGAAGAGAGAGTGGCCCTGGAT 492	
DB	213732	CGGAAAGAGTCTTCACAGAGCTCCGCCAGCTTCGGAAGAGAGAGTGGCCCTGGAT 213791	
QY	493	GAGATCATGAGCTCGNAGGGAAGAAAGACCGCAGCAGACCGCTGTTTACAG 552	
DB	213792	GAGATCATGAGCTCGNAGGGAAGAAAGACCGCAGCAGACCGCTGTTTACAG 213851	
QY	553	CCGGGATCGTTGTGAAAAATATAACGAAGAGCTTTGGGAGAGAAATATCACAGAAGAA 612	
DB	213852	CCGGGATCGTTGTGAAAAATATAACGAAGAGCTTTGGGAGAGAAATATCACAGAAGAA 213911	
QY	613	GGGTCGTTAAGGAAGTATTGACAGGTACACAGCTGTGGTAAAGATGACTGACTCTGGA 672	
DB	213912	GGGTCGTTAAGGAAGTATTGACAGGTACACAGCTGTGGTAAAGATGACTGACTCTGGA 213971	

COMMENT	Contact: MGC help desk Email: cgabs-r@mail.nih.gov Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNL) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: http://www.nisc.nih.gov/ Contact: nisc_mgc@hghri.nih.gov Akter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P., Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Loric,P., Legaspi,R., Maduro,Q.L., Masello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W., Tsugeon,C., Vogt,J.H., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.	
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LILNL at: http://image.llnl.gov Series: IRAK Plate: 123 Row: d Column: 16 This clone was selected for full length sequencing because it passed the following selection criteria: GenomeScan gene prediction, Similarity but not identity to protein.	
source	Location/Qualifiers i. .2710 /organism="Mus musculus" /mol_type="mRNA" /strain="mix FVB/N, C57BL/6J" /db_xref="taxon:10090" /clone="IMAGE:5026202" /issue_type="Mammary tumor. WAP-TGF alpha model. 7 months old, gross tissue." /clone_lib="NCI CGAP_Mam5" /lab_host="DH10B" /note="Vector: pCMV-SPORT6"	
ORIGIN	Query Match 55.5%; Score 611.8; DB 9; Length 2710; Best Local Similarity 98.1%; Pred. No. 1.5e-106; Matches 619; Conservative 0; Mismatches 12; Indels 0; Gaps 0;	
QY	388	GGTACCGTGCCTGAGCTGCTGGGAGCGGAGCATCCGGAAACGGAAGCTTCA 447
DB	2078	GGACCAAGTGCACTGAAGCTGCTGGGGAGCGCAGCATCCGGAAACGGAAGCTTCA 2137
QY	448	CAGAGCTCCGCCAGCTCGGAAGAAAGAAAGTCCGCCCTCGATGAGATCATGAGCTC 507
DB	2138	CAGAGCTCCGCCAGCTCGGAAGAAAGAAAGTCCGCCCTCGATGAGATCATGAGCTC 2197
QY	508	GAAGAGGAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 567
DB	2198	GAAGAGGAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2257
QY	568	AAAAATATACGAGAGCTTGGGAGAAATATACAGAAAGAAAGGCTCGTTAAGGAA 627
DB	2258	AAAAATATACGAGAGCTTGGGAGAAATATACAGAAAGAAAGGCTCGTTAAGGAA 2317
QY	628	GTGATTGACAGGTACACAGCTGTGGTAAAGATGACTCTCTGGAGACAGGCTGAAACTG 687
DB	2318	GTGATTGACAGGTACACAGCTGTGGTAAAGATGACTCTCTGGAGACAGGCTGAAACTG 2377
QY	688	GACGAGCTCATTTAGAGACAGTCAATTCGGCCCCGGGAAAAAGGTTCTAGTTTAAAT 747
DB	2378	GACGAGCTCATTTAGAGACAGTCAATTCGGCCCCGGGAAAAAGGTTCTAGTTTAAAT 2437
QY	748	GGAGGCTACAGAGAAATGAAGCACTTCGATCCATCAATGAGAGGCTTTTCAGCC 807
DB	2438	GGAGGCTACAGAGAAATGAAGCACTTCGATCCATCAATGAGAGGCTTTTCAGCC 2497
QY	808	ACGATAGTCATTTGAACTCGACTTTTGAAGGACGAGAGTTGAAGGTATTCATATGAA 867
Db	2498	ACGATAGTCATTTGAACTCGACTTTTGAAGGACGAGAGTTGAAGGTATTCATATGAA 2557
QY	868	GACATATCTAAACTTGTGCTTGAAGTTTGAAGTTTGAAGTTTGAAGTTTGAAGTT 927
DB	2558	GACATATCTAAACTTGTGCTTGAAGTTTGAAGTTTGAAGTTTGAAGTTTGAAGTT 2617
QY	928	ATCAAAATTTGGTGTAGCCCAAGCACTGTGTAACCTCTACTGTGTAGGGGATTTGTTTGT 987
DB	2618	ATCAAAATTTGGTGTAGCCCAAGCACTGTGTAACCTCTACTGTGTAGGGGATTTGTTTGT 2677
QY	988	ATTAATAAAAAAATAATCATCTATTATAATA 1018
DB	2678	ATTAATAAAAAAATAATCATCTATTATAATA 2708
RESULT 11	AC141995/1 GI:29135466	
LOCUS	Rattus norvegicus (Norway rat)	
DEFINITION	Rattus norvegicus	
ACCESSION	AC141995	
VERSION	AC141995.1	
KEYWORDS	HTG; HTGS PHASE1	
SOURCE	Rattus norvegicus	
ORGANISM	Rattus norvegicus	
REFERENCE	1 (bases 1 to 144118)	
AUTHORS	Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alebrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Balandanaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Devilla,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpthy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuheva,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhinney,S., McLeod,M., McNeill,T., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokemele,O., Okwuonu,G., Olanunsgoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu.L.-L., Puaro,W., Quroz,J., Rachlin,E., Reaves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rivers,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,O., Steimle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,P., Williams,G., Willson,R., Wleczyk,R.,	

```

Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,
Yen,J., Yoon,J., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
Holt,R.A., Smith,H.O., Weinstein,G. and Gibbs,R.A.
Direct Submission
2 (bases 1 to 144118)
Worley,K.C.
Direct Submission
Submitted (21-MAR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GRN
Center clone name: CH230-293H7
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 108995 bases at least Q40
Consensus quality: 114645 bases at least Q30
Consensus quality: 119503 bases at least Q20
Estimated insert size: 114401; sum-of-contigs estimation
Quality coverage: 2x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 48 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1394: contig of 1394 bp in length
* 1395: gap of unknown length
* 1495: contig of 1356 bp in length
* 2850: gap of unknown length
* 2950: gap of unknown length
* 3972: contig of 1022 bp in length
* 4072: gap of unknown length
* 5309: contig of 1237 bp in length
* 5409: gap of unknown length
* 5510: contig of 1155 bp in length
* 6564: gap of unknown length
* 6665: contig of 1785 bp in length
* 8450: gap of unknown length
* 8550: contig of 1354 bp in length
* 9904: gap of unknown length
* 10004: contig of 1743 bp in length
* 11747: gap of unknown length
* 13518: contig of 1672 bp in length
* 13618: gap of unknown length
* 13619: contig of 1352 bp in length
* 14970: gap of unknown length
* 15070: gap of unknown length
* 16373: contig of 1303 bp in length
* 16473: gap of unknown length
* 18359: contig of 1886 bp in length
* 18459: gap of unknown length
* 18460: contig of 1362 bp in length
* 19821: gap of unknown length
* 19922: contig of 1528 bp in length
* 21549: gap of unknown length
* 23269: contig of 1720 bp in length
* 23370: gap of unknown length
* 23770: contig of 2341 bp in length
* 25711: gap of unknown length
* 25810: gap of unknown length
* 27760: contig of 1950 bp in length
*
* 27761: gap of unknown length
* 27860: contig of 1955 bp in length
* 29815: gap of unknown length
* 31213: contig of 1298 bp in length
* 31313: gap of unknown length
* 33042: contig of 1729 bp in length
* 33142: gap of unknown length
* 34194: contig of 1052 bp in length
* 34294: gap of unknown length
* 36020: contig of 1726 bp in length
* 36200: gap of unknown length
* 36220: gap of unknown length
* 37671: contig of 1551 bp in length
* 37771: gap of unknown length
* 39759: contig of 1988 bp in length
* 39859: gap of unknown length
* 42442: contig of 2583 bp in length
* 42542: gap of unknown length
* 44482: contig of 1940 bp in length
* 44582: gap of unknown length
* 46351: contig of 1769 bp in length
* 46451: gap of unknown length
* 48712: contig of 2261 bp in length
* 48812: gap of unknown length
* 51252: contig of 2440 bp in length
* 51352: gap of unknown length
* 54232: contig of 2880 bp in length
* 54332: gap of unknown length
* 57136: contig of 2804 bp in length
* 57236: gap of unknown length
* 59833: contig of 2597 bp in length
* 59933: gap of unknown length
* 62619: contig of 2686 bp in length
* 62719: gap of unknown length
* 65087: contig of 2368 bp in length
* 65187: gap of unknown length
* 68329: contig of 3142 bp in length
* 68429: gap of unknown length
* 72825: contig of 4396 bp in length
* 72925: gap of unknown length
* 76136: contig of 3271 bp in length
* 76236: gap of unknown length
* 76297: contig of 4031 bp in length
* 80328: gap of unknown length
* 83596: contig of 3169 bp in length
* 83696: gap of unknown length
* 88053: contig of 4357 bp in length
* 88153: gap of unknown length
* 91623: contig of 3470 bp in length
* 91723: gap of unknown length
* 96501: contig of 4778 bp in length
* 96601: gap of unknown length
* 101193: contig of 4592 bp in length
* 101293: gap of unknown length
* 106768: contig of 5475 bp in length
* 106868: gap of unknown length
* 111636: contig of 4768 bp in length
* 111736: gap of unknown length
* 117842: contig of 6106 bp in length
* 117942: gap of unknown length
* 117943: contig of 12282 bp in length
* 130324: gap of unknown length
* 130325: contig of 13794 bp in length.
*
* Location/Qualifiers
* 1. 144118
* /organism="Rattus norvegicus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10116"
* /clone="CH230-293H7"
* 1395..1494
* /estimated_length=unknown
* 2851..2950
*
* 48.5%; Score 535; DB 14; Length 144118;

```

Query Match


```
/note="assembly_fragment
clone_end:SP6
vector_side:left"
39773..39872
/estimated_length=unknown
39873..41564
/note="assembly_fragment"
41565..41664
/estimated_length=unknown
41665..47820
/note="assembly_fragment"
47821..47920
/estimated_length=unknown
47921..80877
/note="assembly_fragment"
80878..80977
/estimated_length=unknown
80978..106562
/note="assembly_fragment"
106563..106662
/estimated_length=unknown
106663..130262
/note="assembly_fragment
clone_end:T7
vector_side:right"

ORIGIN
Query Match 46.3%; Score 510.2; DB 14; Length 130262;
Best Local Similarity 85.2%; Pred. No. 2.3e-87;
Matches 609; Conservative 0; Mismatches 93; Indels 13; Gaps 3;

Qy 388 GGTACAGTGCTGCTGAAGCTGCTGGGAGCGCAGCATCCCGGAAACGGAAGTCTTCA 447
Db 100396 GGACCAAGTGCTGCTGCTGAGGAGCAGCATCCCGGAAATGGAAGGCTCTTCA 100455

Qy 448 CAGAGCTCCGCCAGCTGC--GAAGAAGAAGTCCGCCCTGGATGATGATCATGGAG 504
Db 100456 CAGAGCTCTCCAGCTGCTGCAAGAAGAAGATGCTGCCCTGGATGATGATATAGAG 100515

Qy 505 CTCGAAGAGGAAAGAAAGGACCGCACAGACGCTGGTTACAGCCGGGATCGTT 564
Db 100516 ATTAAGAAGAAAGAAAGGACACACAGGAGACTCTGGTTACAGCCGGGAATCAT 100575

Qy 565 GTGAAGATTATACGAAGAAGCTTGGGAGAAATATCACAAGAGAAGGGGCTGTTAG 624
Db 100576 GTGAAGCGTATAGGAAGAAGCTTGGGAGAAATATCAAGAAAGAAAGGGGCTCAT 100635

Qy 625 GAAGTGATTGACAGGTACACAGCTGTGTTAAGATGACTGACTCTGGAGACAGGCTGAAA 684
Db 100636 GAAGAGATTGATGTATACAGCTGTGTTAAGGTGACTGACTCTGGAGACAGGCTGAAA 100695

Qy 685 CTGGACAGAGCTCATTTAGAGACAGTCAATCCGCCCGCGGAAAGGGTTCTAGTTTAA 744
Db 100696 CTGGACAGAGCTCATTTAGAGACAGTCAATCCGCCCGCGGAAAGGGTTCTAGTTTAA 100755

Qy 745 AATGAGGCTACAGAGAAATGAAGGCTCTCAATCCATCATGAGAGGCTTTTCA 804
Db 100756 AATGAGGCTACAGAGAAATGAAGGCAACCTCAATATACATCAATGAGAGGCTTTTTCG 100815

Qy 805 GCCAGTAGTTCATTGAACTGGACCTTTGAAAGAGGACAGCTTGAAGGTATTCAATAT 864
Db 100816 GCTAAGATAGTTCATTGAAATGGACCTTTGAAAGACACAGAGTAGAGCTATTCAATAT 100875

Qy 865 GAAGACATATCTAACTTGTGAGTTTGAAGAAATTTGATAACACACATTTGAAAGTGTGA 924
Db 100876 AAAGACATATCTAATCTTGTGAGTTTGAAGAAATTTGTTAGAAACACATTTGAAATTTGAA 100935

Qy 925 AGCATCAATTTGTTAGTCCAGGCACTGTGTAATCTACTCTGTTAGGAGATTTGTTT 984
Db 100936 AGCATCAATTTGTTAGTCCAGGCACTTTGTAATCTACTCTGTTAGGAGATTTGTTT 100995

Qy 985 TGTATTAAGAAAGAAAGAAATCATCTATTATTAATACTAGTGAATAGTTGGGTAAATTTATA 1044
Db 100996 TGTATTAAGAAAGAAAGAAATCATCTATTATTAATACTAGTGAATAGTTGGGTAAATTTATA 10444
```

```
Db 100996 TGTAT-----ATAAAAAATCATCTATTAAAGTAATACTAAATAATTCGTTAA-----ATA 101045
Qy 1045 ATAAATCTATGTTTTTTTTTAAGTGTAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1099
Db 101046 AAATCTTATGTTTTTTTTTAAGTGTAAAAAAGAAAAAATTAACAGATCTCTACTACAAA 101100

RESULT 13
AC140205 140215 bp DNA linear ROD 29-MAY-2004
LOCUS Mus musculus BAC clone RP24-344P2 from chromosome 8, complete
DEFINITION sequence.
ACCESSION AC140205
VERSION AC140205.2 GI:47131361
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 140215)
AUTHORS Tomlinson, C., Bielicki, L., Cordum, H. and Haakenson, W.
TITLE The sequence of Mus musculus BAC clone RP24-344P2
JOURNAL Unpublished (2001)
REFERENCE 2 (bases 1 to 140215)
AUTHORS McPherson, J. D. and Waterston, R. H.
TITLE Direct Submission
JOURNAL Submitted (23-FEB-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 140215)
AUTHORS Wilson, R. K.
TITLE Direct Submission
JOURNAL Submitted (12-MAY-2004) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 140215)
AUTHORS Wilson, R. K.
TITLE Direct Submission
JOURNAL Submitted (29-MAY-2004) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On May 12, 2004 this sequence version replaced gi:28475441.
----- Genome Center
Center: Washington University Genome Sequencing Center
Web site: http://genome.wustl.edu
Contact: submissions@wustl.edu
----- Summary Statistics
-----
Center project name: M_BB0344P02
-----
```

NOTICE:

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:

The RPCI-24 BAC Library has been constructed by Pieter de Jong and
coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
and/or brain genomic DNA. The clone and detailed information can be
obtained from Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

FEATURES

source
Location/Qualifiers
1. .140215
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="8"
/map="8"
/clone="RP24-344F2"
/clone.lib="RPC1-24"
32. .81
/rpt_family="ERVK"
333. .426
/rpt_family="ERVK"
432. .600
/rpt_family="ERVK"
605. .763
/rpt_family="ERVK"
830. .1157
/rpt_family="ERVK"
1158. .1387
/rpt_family="ERVK"
1426. .1641
/rpt_family="ERVK"
1657. .1779
/rpt_family="ERVK"
2071. .2147
/rpt_family="ERVK"
2242. .2391
/rpt_family="ERVK"
2832. .2881
/rpt_family="ERVK"
3133. .3226
/rpt_family="ERVK"
3232. .3400
/rpt_family="ERVK"
3405. .3563
/rpt_family="ERVK"
3630. .3957
/rpt_family="ERVK"
3958. .4187
/rpt_family="ERVK"
4226. .4441
/rpt_family="ERVK"
4457. .4579
/rpt_family="ERVK"
4871. .4947
/rpt_family="ERVK"
5042. .5191
/rpt_family="ERVK"
5632. .5681
/rpt_family="ERVK"
5933. .6026
/rpt_family="ERVK"
6032. .6200
/rpt_family="ERVK"
6205. .6363
/rpt_family="ERVK"
6430. .6757
/rpt_family="ERVK"
8931. .9312
/rpt_family="L1"
9434. .9660
/rpt_family="L1"
9692. .9931
/rpt_family="L1"
9959. .10098
/rpt_family="L1"
10770. .10883
/rpt_family="L1"
11090. .11220
/rpt_family="Alu"
11240. .11299

repeat_region
/rpt_family="L1"
11704. .11890
/rpt_family="ERV1"
11977. .12288
/rpt_family="ERV1"
13460. .13519
/rpt_family="ERV1"
13586. .13801
/rpt_family="RMER10B"
18269. .18413
/rpt_family="RMER10B"
19731. .21946
/rpt_family="ERV1"
21953. .23462
/rpt_family="ERV1"
23875. .24058
/rpt_family="ERV1"
24558. .24711
/rpt_family="ERVK"
25065. .25121
/rpt_family="L1"
26157. .26279
/rpt_family="RMER10A"
27231. .27425
/rpt_family="RMER10A"
27505. .27718
/rpt_family="RMER10A"
27956. .28084
/rpt_family="MaLR"
28245. .28470
/rpt_family="MaLR"
30676. .30798
/rpt_family="MaLR"
31586. .31640
/rpt_family="L1"
31634. .32651
/rpt_family="L1"
33734. .33932
/rpt_family="MERL_type"
35881. .36272
/rpt_family="MaLR"
38599. .39351
/rpt_family="L1"
40422. .40554
/rpt_family="L1"
41164. .41266
/rpt_family="B2"
42685. .43336
/rpt_family="L1"
45098. .45157
/rpt_family="ERV1"
45351. .45480
/rpt_family="ERV1"
46796. .47083
/rpt_family="L1"
47434. .47573
/rpt_family="B2"
47713. .48088
/rpt_family="L1"
48495. .48643
/rpt_family="Alu"
49036. .49254
/rpt_family="MERL_type"

Query Match 46.3%; Score 510.2; DB 9; Length 140215;
Best Local Similarity 85.2%; Pred No. 2.2e-87;
Matches 609; Conservative 0; Mismatches 93; Indels 13; Gaps 3;
Qy 388 GGTACCACTGCACCTGAAGCTGCTGGGAGCGCAGCATCCGGGAAACGGAAAGAGTCTTCA 447
Db 82385 GGACCAAGTGCACCTGAAGCTGCTGAGGAGCACAGCATCCGGGAAATGGAAAGGCTCTTCA 82444
Qy 448 CAGAGCTCGGCCACCTGCG---GNAGAGAGAGAGTGGCCCTGGATCAGATCATGGAG 504

||||| 1146 AATGAGAAGACTTTTTCAGCTACTATCGTCATTGAAACTGGCCCTTTAAAGGAGCGAGA 1205
QY 847 GTTGAAGGATTTCAATATGAAGACATATCTAAACTTGTCTGAGTTTGAAGAAATTTGATAAC 906
Db 1206 GTTGAAGGAATTCATATGAAGACATTTCTAAACTTGCTGAGTTTGAAGAAATTTGTTAAC 1265
QY 907 AACACATTGAACCTGTGAAGCATCAAAATGGTGGTTAGCCCAAGGCATGTGTAACCTCTACT 966
Db 1266 AATACATTAAAATCTTAAAGCATCAAAATGGTGGTTAGCCCAAGGCATTTATGAGACTCTACT 1325
QY 967 GTGTTAGGGGATTTCTTTGTATTAAATAAATAAATAAATCATCTATTAAATACTAGTGA 1026
Db 1326 GTGTTAGGGATATTCTTTTGTATTAAATAAATAAATCATCTATTAAATACTAGTGA 1385
QY 1027 TAGTTGGGTAATTTATATAAATACTATGTTTTTTTTTAAAGTGTAATAAAA 1077
Db 1386 TGTTGAGCTAATCTTGAGAGAAATTAATTATGCTCATGAGGTATCAAA 1436

RESULT 15
HSAJ5273
LOCUS HSAJ5273 1528 bp mRNA linear PRI 13-DEC-2000
DEFINITION Homo sapiens mRNA for Kin17 protein.
ACCESSION AJ005273
VERSION AJ005273.1 GI:3850703
KEYWORDS Kin17 gene.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1
Kannouche, P., Mauffrey, P., Pinon-Lataillade, G., Mattei, M.G.,
Saxasin, A., Daya-Grosjean, L. and Angulo, J.F.
Molecular cloning and characterization of the human KIN17 cDNA
encoding a component of the UVC response that is conserved among
metazoans
Carcinogenesis 21 (9), 1701-1710 (2000)
10964102
2
Mauffrey, P.
Direct Submission
Submitted (08-Apr-1998) Mauffrey P., LGR, CEA-DSV-DRR, BP 6 -
Fontenay aux Roses Cedex, 92265, FRANCE
Revised by [3]
3 (bases 1 to 1528)
Mauffrey, P.
Direct Submission
Submitted (07-OCT-1998) Mauffrey P., LGR, CEA-DSV-DRR, BP 6 -
Fontenay aux Roses Cedex, 92265, FRANCE
On Nov 7, 1998 this sequence version replaced gl:3046739.
Kannouche P., Mauffrey P., Pinon-Lataillade G., Biard D.S.F.,
Angulo J.F.
Sequences codant pour la Proteine kin17 et leurs Applications -
French Patent Nr 97 15536 1:1-50(1997).
Location/Qualifiers
1..1528
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/sex="male"
/tissue type="testis"
1..1528
/gene="Kin17"
66..1247
/gene="kin17"
/codon_start=1
/protein_id="CAA06462.1"
/db_xref="GI:3850704"
/db_xref="GOA:O60870"
/db_xref="UniProt/TrEMBL:O60870"
/translation="MGKSDFLTPKAIANRIKSKGLQKLRWYQMCKQCRDENGPKCH

CMSEHQORLLASENPQPMFYSESEFRNDFLELLRRFGTKRVHNNIVVNEYISHR
EHIHMNATQWETLDTFTKMLRGELCKVDETPKGWIQYIDRDPETIRRLQLEKKK
QDLDDDEKTAKEFIEBQVRRGLEKREQVPTFTELSRENDEKVTNLSKGACSSGAT
SSKSTLSPSALKITIGSSASVSRKESQSSQTSKEKKKKSALDIMEIBEKSKRTAR
TDYWLQPIIVKIIITKLGKEXYHKKAIVKEVIDKYTAVVRMIDSGDKLKDQTHLET
VIPAPGRKILVLNGYGRNEGTLESINEKTFSATIVIETGPLKGRRVEGIQYEDISKL
A"

ORIGIN
Query Match 43.1%; Score 475; DB 8; Length 1528;
Best Local Similarity 65.4%; Pred. No. 2e-80;
Matches 897; Conservative 0; Mismatches 180; Indels 294; Gaps 2;

QY 1 ATGGGCAAGTCGATTTCTGAGCCCAAGGCATCGCAATAGAAATTAAGTCCAAAGG 60
Db 66 ATGGGAAAGTCGATTTCTTACTCCCAAGGCTATCGCCAACAGGATCAAGTCCAAGGG 125
QY 61 CTCAGAGAAGCTTCGCTGGTACTGCCAGATGTGCCAAAAGCAATGCCGACAGAAATGCG 120
Db 126 CTCAGAGAAGCTACGCTGGTATTGCCAGATGTGCCAGAAGCAGTGCCTGGGACGAGAAATG 185
QY 121 TTAAAGTCTCATCTGATGCTGAATCTCATCAAGACAACTGTTGCTGGCTTCAGAAAC 180
Db 186 TTAAAGTCTCATCTGATGCTCGAATCTCATCAGACAACTATTGCTGGCTTCAGAAAT 245
QY 181 CCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTCGAAATGACTTTCTGAACTTCTG 240
Db 246 CCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTTCCGAAATGACTTTCTGAACTTCTC 305
QY 241 AGGCGACGCTTTGGCAGCTAAAGGGTCCCAACAACTTGTCTACAATGAATACATCAGC 300
Db 306 AGGAGACGCTTTGGCAGCTAAAGGGTCCCAACAACTTGTCTACAATGAATACATCAGC 365
QY 301 CACCGAGAGCAGATCCACATGAACGCTACCCAGTGGGAGACACACCGACCTTTTACCAG 360
Db 366 CACCGAGAGCAGATCCACATGAATGCCACTCAGTGGGAACTCTGACTGATTTTACTAAG 425
QY 361 TGGCTGGGACAGAGAGGGCTTGTGTAA----- 387
Db 426 TGGCTGGGACAGAGAGGGCTTGTGTAAAGTGGACGAGACACCAAAAGGTGGTATATTTCAG 485
QY 388 ----- 387
Db 486 TACATAGACAGGACCCAGAAACTATATCCCGCGGCAACTGGAACTGGAGAAAGAAAGAA 545
QY 388 ----- 387
Db 546 CAGGACCTTGATGATGAAGAAAAAACTGCCAAATTTATTGAAGAGCAAGTGAGAGAGGC 605
QY 388 ----- 387
Db 606 CTGGAAGGAAAGGAACAGGAGGTCCCTACTTTTACGGAATTAAGCAGAGAAATGATGAA 665
QY 388 ----- 387
Db 666 GAGAAAGTCAGCTTTAAATTTGAGTAAAGAGCATGTAGCTATCCGGAGCAACATCTTCC 725
QY 388 -----GGTACCAGTGCCTGAAAGTCTGGGAGCGCAGCATCCCGGAAA 432
Db 726 AAGTCAAGTACTCTCGGACCGAGTGCACTGAAGACGATAGGAAGTTTCAGATCAGTGA 785
QY 433 CGGAAAGAGTCTTTCACAGAGTCCGCCAGCCTGC-----GAAGAAGNAGAAAGTCCGCC 486
Db 786 CGAAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAAAGAAAAAGAAATCTGCA 845
QY 487 CTGGATGAGATCATCGAGCTCGAAGAGGAAAAAGAAAGCAAGCAGACACGCTCGG 546
Db 846 CTGGATGAAATCATCGAGATTTGAAGAGGAAAAAGAAAGAACTGCCCGAACACACTACTG 905
QY 547 TTACAGCCGGGATCGTTGTGAAAAATTAACAGAAAGAGCTTTGGGAGAAAAATATCA 606
Db 906 CTACAGCCTGAAATTTATTGTGAAAAATTTATACCAAGAAACTGGGAGAGAAATATCATAG 965

This Page Blank (uspic)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: November 26, 2005, 12:48:57 ; Search time 706.843 Seconds
(without alignments)
10390.557 Million cell updates/sec

Title: US-09-555-529-2
Perfect score: 1102
Sequence: 1 atggcgagctcggtttctt.....aaaaaaaaaaaaaaaaaaaaa 1102

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues 9993994

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : N Geneseq_21.*
- 1: Geneseqn1980s.*
 - 2: Geneseqn1990s.*
 - 3: Geneseqn2000s.*
 - 4: Geneseqn2001s.*
 - 5: Geneseqn2001bs.*
 - 6: Geneseqn2002as.*
 - 7: Geneseqn2002bs.*
 - 8: Geneseqn2003as.*
 - 9: Geneseqn2003bs.*
 - 10: Geneseqn2003cs.*
 - 11: Geneseqn2003ds.*
 - 12: Geneseqn2004as.*
 - 13: Geneseqn2004bs.*
 - 14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1102	100.0	1102	2	AX85550 CDNA of a
2	799.2	72.5	1390	2	AX85570 CDNA of a
3	790	71.7	1458	2	AQ79936 Murine Ki
4	704.6	63.9	1002	2	AX85551 CDNA of a
5	422.2	38.3	1296	2	AX85549 CDNA of a
6	385.4	35.0	679	13	ADQ56943 Novel can
7	329.2	29.9	1002	2	AX85552 Probe der
8	269.2	24.4	591	4	AH34095 Human col
9	190	17.2	1372	12	ADP22451 Sea-equir
10	174	15.8	1459	3	AAC47248 Arabidops
11	168.8	15.3	1614	13	ADO84255 Plant ful
12	168.8	15.3	1713	13	ADX47874
13	168.8	15.3	1764	3	AAAS5589
14	167.6	15.2	1713	13	ADT17629
15	167.6	15.2	1780	3	AAAS5291
16	167.6	15.2	1780	3	AAAS5290 CDNA enco
17	163	14.8	1241	4	ABL29199 Drosophil
18	163	14.8	3313	4	ABL29198 Drosophil
19	163	14.8	4394	4	ABL29184 Drosophil

20	163	14.8	4582	4	ABL18194	Ab118194 Drosophil	
21	136	12.3	1032	3	AAC48794	Aac48794 Arabidops	
22	130.8	11.9	549	3	AAC94929	Aac94929 Cat flea	
C	23	116.8	10.6	398	13	ACF87901	Acf87901 Human SIR
24	95.4	8.7	300	3	AAA01559	Aaa01559 Human col	
25	93.6	8.5	270	6	ABL73715	Ab173715 Corn tass	
26	85	7.7	696	3	AAC38477	Aac38477 Arabidops	
27	67.4	6.1	431	3	AAC48801	Aac48801 Arabidops	
C	28	66.2	6.0	375	5	ABV44911	Abv44911 Human pro
29	65	5.9	65	6	ABN51894	Abn51894 Mouse spl	
C	30	63	5.7	396	13	ACN63020	Acn63020 Cotton ca
31	62.6	5.7	552	13	ACN45312	Acn45312 Cotton pr	
C	32	61.2	5.6	723	5	ADI71984	Adi71984 Human ova
C	33	61.2	5.6	723	5	ADL37134	Adl37134 Human ova
C	34	59.4	5.4	390	5	ADL43918	Adl43918 Human ova
C	35	59.2	5.4	337	5	ABV08167	Abv08167 Human pro
	36	59	5.4	2229	2	AAQ23462	Aaq23462 Human B-r
	37	59	5.4	2229	10	AAQ64082	Aaq64082 DNA #3 re
	38	59	5.4	2229	12	ADL16255	Adl16255 Human B-r
	39	59	5.4	2510	2	AAx21931	Aax21931 Human B-r
	40	59	5.4	2510	6	ABK84128	Abk84128 Human CDN
	41	59	5.4	2510	6	ABL57049	Ab157049 Human pro
	42	59	5.4	2510	6	AAU44844	Aad44844 Human B-r
	43	59	5.4	2510	9	ADA02937	Ada02937 Human BRA
	44	59	5.4	2510	10	ADB72675	Adb72675 Human BRA
	45	59	5.4	2510	10	ADC85417	Adc85417 Human Bra

ALIGNMENTS

RESULT 1
AX85550
ID AX85550 standard; cDNA; 1102 BP.
XX
AC AX85550;
XX
DT 07-SEP-1999 (first entry)
XX
DE cDNA of a gene coding for a mouse deleted kin17 protein.
XX
KW kin17 protein; cell proliferation; fertility; hyperproliferative disease;
KW protein interaction; curved DNA; HIV replication; HIV integration;
KW repair enzyme; ss.
XX
OS Mus sp.
XX
PN FR2772046-A1.
XX
PD 11-JUN-1999.
XX
PF 09-DEC-1997; 97FR-00015536.
XX
PR 09-DEC-1997; 97FR-00015536.
XX
(COMS) COMMISSARIAT ENERGIE ATOMIQUE.
XX
PI Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
DR WPI; 1999-359999/31.
XX
PT New DNA coding for human kin17 protein - useful for controlling cell
PT proliferation or fertility.
XX
PS Claim 4; Page 31; 69pp; French.
XX
CC The present sequence encodes a mouse kin17 protein with amino acids 129-
CC 228 deleted. The mammalian kin17 protein is useful for preparing a
CC medicament for controlling cell proliferation or for controlling
CC fertility. The medicaments can also be used to treat hyperproliferative
CC diseases. Fragments between amino acids 55 and 235 (preferably between
CC amino acids 129 and 228) of a mammalian kin17 protein are useful for
CC regulating the interaction between proteins and curved DNA. The fragment

CC can be used to block replication of HIV or its integration into the human
CC genome or to target repair enzymes to curved DNA sites. Expression
XX vectors for kin17 can be used for controlling cell proliferation

SQ Sequence 1102 BP; 373 A; 205 C; 272 G; 252 T; 0 U; 0 Other;
Query Match 100.0%; Score 1102; DB 2; Length 1102;
Best Local Similarity 100.0%; Pred. No. 1.5e-208;
Matches 1102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCCAGTCCGATTTCTGAGCCCAAGGCCATCGCCCAATAGATTAAAGTCCAAAGG 60
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1 ATGGCCAGTCCGATTTCTGAGCCCAAGGCCATCGCCCAATAGATTAAAGTCCAAAGG 60
QY 61 CTCGAGAGCTTCGCTGGTACTGCCAGATGTGCCAAAAGCAATGCCGCGACGAGAATGCG 120
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
61 CTCGAGAGCTTCGCTGGTACTGCCAGATGTGCCAAAAGCAATGCCGCGACGAGAATGCG 120
QY 121 TTTAAGTGTCTATGTCTGAATCTCATCAAAAGCAAACTGTGTGCTGCTTCAGAAAAC 180
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
121 TTTAAGTGTCTATGTCTGAATCTCATCAAAAGCAAACTGTGTGCTGCTTCAGAAAAC 180
QY 181 CCTCAGCAGTTCATGATTTTTCAGAGGAATCCGAAATGACTTTCTGGAATCTTG 240
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
181 CCTCAGCAGTTCATGATTTTTCAGAGGAATCCGAAATGACTTTCTGGAATCTTG 240
QY 241 AGCGCAGCTTTGGCACTAAAAGGGTCCACAAACATTTGCTCAATGAATACATCAGC 300
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
241 AGCGCAGCTTTGGCACTAAAAGGGTCCACAAACATTTGCTCAATGAATACATCAGC 300
QY 301 CACCGAGACATCCACATGAACGCTACCCAGTGGGAGACACTGACCGACTTTTACCAG 360
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
301 CACCGAGACATCCACATGAACGCTACCCAGTGGGAGACACTGACCGACTTTTACCAG 360
QY 361 TGGCTGGGAGAGAGGGCTTGTAAAGTACAGTCCAGTCAAGCTGCTGGGGAGCGCA 420
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
361 TGGCTGGGAGAGAGGGCTTGTAAAGTACAGTCCAGTCAAGCTGCTGGGGAGCGCA 420
QY 421 GCATCCGGGAAACGGAAGAGTCTTCACAGAGCTCCGCCAGCTCGAAGAGAGAGAG 480
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
421 GCATCCGGGAAACGGAAGAGTCTTCACAGAGCTCCGCCAGCTCGAAGAGAGAGAG 480
QY 481 TCGGCTCGATGAGATCATGAGCTCGAAGAGGAAAGAAAGGACCGCACGACAGAC 540
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
481 TCGGCTCGATGAGATCATGAGCTCGAAGAGGAAAGAAAGGACCGCACGACAGAC 540
QY 541 GCTGCTTACCGCGGGGATCGTTGAAATTTATAACGAAGAGCTTCGGGAGAAATAT 600
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
541 GCTGCTTACCGCGGGGATCGTTGAAATTTATAACGAAGAGCTTCGGGAGAAATAT 600
QY 601 CACAAGAAAGGGTCTGTTAAGGAAGTGTATGACAGGTACACAGCTGTGGTAAAGATG 660
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
601 CACAAGAAAGGGTCTGTTAAGGAAGTGTATGACAGGTACACAGCTGTGGTAAAGATG 660
QY 661 ACTGACTCTGGAGACAGGCTGAAATGACACGACTCAITTTAGACAGCTCAITTCGGGC 720
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
661 ACTGACTCTGGAGACAGGCTGAAATGACACGACTCAITTTAGACAGCTCAITTCGGGC 720
QY 721 CCGGGGAAAGGGTCTAGTTTAAATGGAGGCTACAGAGGAATGAAGGCACTCTCGAA 780
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
721 CCGGGGAAAGGGTCTAGTTTAAATGGAGGCTACAGAGGAATGAAGGCACTCTCGAA 780
QY 781 TCCATCAATGAGAAGGCTTTTCAGGCCAGATAGTCAITTAAGTCAAGTCTTGAAGGA 840
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
781 TCCATCAATGAGAAGGCTTTTCAGGCCAGATAGTCAITTAAGTCAAGTCTTGAAGGA 840
QY 841 CGCAGAGTTGAAGGTATTCATATGAAGACATATCTAACTGCTTGAAGTCTGAAATTT 900
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
841 CGCAGAGTTGAAGGTATTCATATGAAGACATATCTAACTGCTTGAAGTCTGAAATTT 900
QY 901 GATACACACATTTGAAGCTGAAGCATCAATTCGTTGTTAGCCCAAGGCACTGTGTAAC 960
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
901 GATACACACATTTGAAGCTGAAGCATCAATTCGTTGTTAGCCCAAGGCACTGTGTAAC 960

QY 961 TCTACTGTGTAGGGATTTGTTTATTAATAAAAAAAAAAATCATCTATTAAATACT 1020
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
961 TCTACTGTGTAGGGATTTGTTTATTAATAAAAAAAAAAATCATCTATTAAATACT 1020
QY 1021 AGTGAATAGTTGGGTAAATTTATATAAATCTATGTTTTTTTAAAGTGTAATAAAAA 1080
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1021 AGTGAATAGTTGGGTAAATTTATATAAATCTATGTTTTTTTAAAGTGTAATAAAAA 1080
QY 1081 AAAAAAAAAAAAAAAAAAAAAA 1102
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1081 AAAAAAAAAAAAAAAAAAAAAA 1102

RESULT 2
AA85570
ID AA85570 standard; cDNA; 1390 BP.

XX AA85570;

XX 07-SEP-1999 (first entry)

XX cDNA of a gene coding for the murine kin17 protein.

DE Mouse; kin17 protein; cell proliferation; fertility;
XX hyperproliferative disease; protein interaction; curved DNA;
KW HIV replication; HIV integration; repair enzyme; ss.

OS Mus sp.

XX FR2772046-A1.

XX 11-JUN-1999.

XX 09-DEC-1997; 97FR-00015536.

XX 09-DEC-1997; 97FR-00015536.

XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

PI Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;

XX WPI; 1999-359999/31.

PT New DNA coding for human kin17 protein - useful for controlling cell proliferation or fertility.

PS Claim 21; Page 35-36; 69pp; French.

CC The present sequence encodes a murine kin17 protein. The mammalian kin17 protein is useful for preparing a medicament for controlling cell proliferation or for controlling fertility. The medicaments can also be used to treat hyperproliferative diseases. Fragments between amino acids 55 and 235 (preferably between amino acids 129 and 228) of a mammalian kin17 protein are useful for regulating the interaction between proteins and curved DNA. The fragment can be used to block replication of HIV or its integration into the human genome or to target repair enzymes to curved DNA sites. Expression vectors for kin17 can be used for controlling cell proliferation

XX Sequence 1390 BP; 482 A; 256 C; 353 G; 299 T; 0 U; 0 Other;

Query Match 72.5%; Score 799.2; DB 2; Length 1390;
Best Local Similarity 79.1%; Pred. No. 1.4e-148;
Matches 1099; Conservative 0; Mismatches 3; Indels 288; Gaps 1;

QY 1 ATGGCCAGTCCGATTTCTGAGCCCAAGGCCATCGCCCAATAGATTAAAGTCCAAAGG 60
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1 ATGGCCAGTCCGATTTCTGAGCCCAAGGCCATCGCCCAATAGATTAAAGTCCAAAGG 60

QY 61 CTCGAGAGCTTCGCTGGTACTGCCAGATGTGCCAAAAGCAATGCCGCGACGAGAATGCG 120
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
61 CTCGAGAGCTTCGCTGGTACTGCCAGATGTGCCAAAAGCAATGCCGCGACGAGAATGCG 120

```
QY 121 TTTAAGTGTCTGATGCTGATCTCATCAAGACAACTGTTGCTGCTTCAGAAAC 180
Db 121 TTTAAGTGTCTGATGCTGATCTCATCAAGACAACTGTTGCTGCTTCAGAAAC 180
QY 181 CCTCAGCAGTTATGATATTTTTCAGAGGAATCCGAAATGACATTTCTGGAACTTCTG 240
Db 181 CCTCAGCAGTTATGATATTTTTCAGAGGAATCCGAAATGACATTTCTGGAACTTCTG 240
QY 241 AGCGGACGCTTTGGCACTAAAGGGTCCACAACAATTGCTCTCAATGAATACATCAGC 300
Db 241 AGCGGACGCTTTGGCACTAAAGGGTCCACAACAATTGCTCTCAATGAATACATCAGC 300
QY 301 CACCGAGAGCAGATCCACATGACGTACCCAGTGGGAGACACTGACCGACTTTTACCAAG 360
Db 301 CACCGAGAGCAGATCCACATGACGTACCCAGTGGGAGACACTGACCGACTTTTACCAAG 360
QY 361 TGGCTGGGACAGAGGGCTTGTGTAAA----- 387
Db 361 TGGCTGGGACAGAGGGCTTGTGTAAA----- 387
QY 388 ----- 387
Db 421 TACATAGACAGACCCAGAAACCATCCGTCGGCACTGGAAATTAGAAAAAAGAAAG 480
QY 388 ----- 387
Db 481 CAAGATCTGACGATGAAGAAAAAACTGCCAAGTTTATTGAGGAGCAGGTGAGAAAGGC 540
QY 388 ----- 387
Db 541 CTGGAAGGGAAAGACGAGGAGACACCTGTTTTTACAGAACTTAGCCGAGAAATGAGGA 600
QY 388 ----- 387
Db 601 GAAAAAGTTACGTTCAATCTGAATTAAGGAGCGGTGGCTCAGCGGAGCTACAACATCC 660
QY 388 -----GATACAGTGCATCAAGTCTGGGAGCGCAGCATCCGGAAA 432
Db 661 AAGTCAAGCTCTTTGGGACCAAGTGCACTGAAGTCTGGGAGCGCAGCATCCGGAAA 720
QY 433 CGGAAGAGTCTTACAGAGCTCGCCAGCTCGGAGGAGGAGAGTCTGGCTGAT 492
Db 721 CGGAAGAGTCTTACAGAGCTCGCCAGCTCGGAGGAGGAGAGTCTGGCTGAT 780
QY 493 GAGATCATGAGGCTCGAAGAGGAAAAAGGACCGCACGAGCAGCGCTGTTACAG 552
Db 781 GAGATCATGAGGCTCGAAGAGGAAAAAGGACCGCACGAGCAGCGCTGTTACAG 840
QY 553 CCGGGGATCGTTGTGAAAATTATACGAAGAGCTTGGGAGAAATATCAAGAAAGAA 612
Db 841 CCGGGGATCGTTGTGAAAATTATACGAAGAGCTTGGGAGAAATATCAAGAAAGAA 900
QY 613 GGGTCTGTAAGGAAGTATTGACAGGTACACAGCTGTGTTAAAGATGACTCTGGA 672
Db 901 GGGTCTGTAAGGAAGTATTGACAGGTACACAGCTGTGTTAAAGATGACTCTGGA 960
QY 673 GACAGGCTGAAACTGGACCCAGACTCATTTAGAGACAGTCAATTCGGCCCGGGGAAAGG 732
Db 961 GACAGGCTGAAACTGGACCCAGACTCATTTAGAGACAGTCAATTCGGCCCGGGGAAAGG 1020
QY 733 GTTCTAGTTTTAAATGAGGGCTACAGAGGAATGAAGGCACTCTCGAATCCATCAATGAG 792
Db 1021 GTTCTAGTTTTAAATGAGGGCTACAGAGGAATGAAGGCACTCTCGAATCCATCAATGAG 1080
QY 793 AAGGCTTTTTCAGCCAGATAGTATTCAAACTCGACTTTGAAAGGAGCGCAGAGTTGAA 852
Db 1081 AAGGCTTTTTCAGCCAGATAGTATTCAAACTCGACTTTGAAAGGAGCGCAGAGTTGAA 1140
QY 853 GGTATTCATATGAAGACATATCTAAACTTCTGAGTTTGAATTTGATAACAACACA 912
Db 1141 GGTATTCATATGAAGACATATCTAAACTTCTGAGTTTGAATTTGATAACAACACA 1200
```

```
QY 913 TTGAAACTGTGAAGCATCAAAATGGTGTGTAGCCAGGCACTGTGTAACCTCTACTGTGTTA 972
Db 1201 TTGAAACTGTGAAGCATCAAAATGGTGTGTAGCCAGGCACTGTGTAACCTCTACTGTGTTA 1260
QY 973 GGGGATTTGTTGTTATTAATAAAAAAATCAATCATCTATTTAAATACTAGTGAATAGTTG 1032
Db 1261 GGGGATTTGTTGTTATTAATAAAAAAATCAATCATCTATTTAAATACTAGTGAATAGTTG 1320
QY 1033 GGTAAATTTATATAAATCTATGTTTTTTTAAAGTGTAAAAAATAAAAAAATAAAAAA 1092
Db 1321 GGTAAATTTATATAAATCTATGTTTTTTTAAAGTGTAAAAAATAAAAAAATAAAAAA 1380
QY 1093 AAAAAAATA 1102
Db 1381 AAAAAAATA 1390
RESULT 3
AAQ79936
ID AAQ79936 standard; cDNA; 1458 BP.
XX
AC AAQ79936;
XX
DT 25-MAR-2003 (revised)
DT 06-SEP-1995 (first entry)
XX
DE Murine Kin17 cDNA.
XX
KW chromosomal rearrangement; kin17 protein; SOS DNA repair system; RecA;
KW genotoxic agent; zinc finger; DNA binding protein; ss.
XX
OS Mus musculus.
XX
FH Key
FH primer_bind Location/Qualifiers
FT /tag= b complement(1..21)
FT /note= "Oligo L (AAQ79946) binds to complementary strand"
FT misc_feature 22..1434
FT /tag= 1
FT /label= kin17_cDNA
FT /note= "nucleotides 1-1414; the genomic DNA contains at least two introns within this sequence, see Comments"
FT primer_bind 32..49
FT /tag= c
FT /note= "Oligo C (AAQ79938) binding site"
FT CDS 46..1221
FT /tag= a
FT /product= "kin17"
FT /note= "N's in the sequence denote illegible residues"
FT primer_bind complement(67..86)
FT /tag= d
FT /note= "Oligo S (AAQ79947) binds to complementary strand"
FT primer_bind 274..297
FT /tag= e
FT /note= "Oligo D (AAQ79939) binding site"
FT primer_bind complement(339..360)
FT /tag= f
FT /note= "Oligo K (AAQ79945) binds to complementary strand"
FT primer_bind 451..474
FT /tag= g
FT /note= "Oligo J (AAQ79944) binding site"
FT primer_bind complement(550..567)
FT /tag= h
FT /note= "Oligo E (AAQ79940) binds to complementary strand"
FT primer_bind 802..825
FT /tag= i
FT /note= "Oligo F (AAQ79941) binding site"
FT primer_bind complement(839..862)
FT /tag= j
FT /note= "Oligo G (AAQ79942) binds to complementary strand"
FT primer_bind complement(1435..1458)
FT /tag= k
FT /note= "Oligo B (AAQ79937) binds to complementary strand"
FT
```

XX FR2706487-A1.
XX PD 23-DEC-1994.
XX PF 15-JUN-1993; 93PR-00007171.
XX PF 15-JUN-1993; 93PR-00007171.
XX PR (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
XX PA Angulo-Mora JF, Tissier A, Prelat G, Mauffrey P, Guilly M;
XX PI WPI; 1995-039031/06.
XX DR Purified murine kin17 protein prepn. for detecting chromosomal
XX PT rearrangements - also related antibodies, human and murine DNA, primers,
XX PT probes and vectors, used to assess damage caused by genotoxic agents.
XX PS Claim 9; Page 33; 54pp; French.
XX CC The murine Kin17 protein includes a zinc finger domain (see AAR66766),
CC recognises single- and double-stranded DNA (partic. regions of secondary
CC structure), has apparent mol. wt. 43 kD and is recognised by both anti-
CC kin17 antibodies and antibodies against the RecA protein of E.coli. The
CC kin17 protein is involved in DNA repair; it can be used to monitor
CC chromosomal rearrangements following exposure to genotoxic agents. The
CC kin17 cDNA sequence AAQ79936 consists of a 1414 nucleotide sequence,
CC flanked by primer binding sites; the genomic kin17 DNA is claimed in
CC which an intron is inserted at position 137 of the 1414 nucleotide cDNA
CC and a second intron is located between nucleotides 339-429 of the cDNA.
XX CC (Updated on 25-MAR-2003 to correct PN field.)
SQ Sequence 1458 BP; 496 A; 272 C; 372 G; 311 T; 0 U; 7 Other;
Query Match 71.7%; Score 790; DB 2; Length 1458;
Best Local Similarity 78.5%; Pred. No. 9.1e-147; Matches 1091; Conservative 0; Mismatches 11; Indels 288; Gaps 1;
1 ATGGGCAAGTCGGATTTCTGAGCCCAAGGCCATCGCCAAATAGAAATTAAGTCCAAAGGG 60
46 ATGGGCAANTCGGATTTCTGAGCCCAAGGCCATCGCCAAATAGAAATTAAGTCCAAAGGG 105
61 CTCGAGAGCTTCGCTGATGTCGAGATGTCGCAAGAGCAATGCGGACGAGAAATGGC 120
106 CTCGAGAGCTTCGCTGATGTCGAGATGTCGCAAGAGCAATGCGGACGAGAAATGGC 165
121 TTAAAGTGTCACTGTATGTCGAATCTCATCAAGACAACTGTGTCGCTTCAGAAAC 180
166 TTAAAGTGTCACTGTATGTCGAATCTCATCAAGACAACTGTGTCGCTTCAGAAAC 225
181 CTCGAGAGCTTCGATTTATTTTTCAGAGGAATTCGGAATGACCTTCGGAATCTCTG 240
226 CTCGAGAGCTTCGATTTATTTTTCAGAGGAATTCGGAATGACCTTCGGAATCTCTG 285
241 AGCGAGCGTTTGGCACTAAAGGGTCCACACCAATGTCCTACAAATGATACATCAGC 300
286 AGCGAGCGTTTGGCACTAAAGGGTCCACACCAATGTCCTACAAATGATACATCAGC 345
301 CACCCAGAGCACATCCACATGAACGCTACCCAGTGGGAGACACTGACCGACTTTTACCAG 360
346 CACCCAGAGCACATCCACATGAACGCTACCCAGTGGGAGACACTGACCGACTTTTACCAG 405
361 TGGCTGGGCGAGAGGGCTTTGTGTAAA----- 387
406 TGGCTGGGCGAGAGGGCTTTGTAAAGTGGATGAGACACCGAAGGCTGGTACATTCAG 465
388 ----- 387
466 TACATAGACAGAGACCCAGAAACCATCCGTCGGCAACTGGAATTAGAAAAAAGAGAAG 525
388 ----- 387

Db 526 CAAGATCTGGACGATGAAGAAAAAATCTCCCAAGTTCAATTGAGGAGCAGGTCAAGAGAGGC 585
QY 388 ----- 387
Db 596 CTNGAAGGGAAGAGAGCAGGAGACACCTGTTTTTACAGAACTTAGCCGAGAAAAATGAGGAA 645
QY 388 ----- 387
Db 646 GAAAAAGTTACGTTCAATCTGAATAAAGAGCGGGTGGCTCAGCGGGAGGCTACAACATCC 705
QY 388 -----GGTACCAAGTGCCTGAAGCTGCTGGGGAGCGCAGCATCCCGGAAA 432
Db 706 AAGTCAAGCTCTTTGGGACCAAGTGCATGAAGCTGCTGGGGAGCGCAGCATCCCGGAAA 765
QY 433 CGGAAAGAGTCTTACAGAGCTCCGCCAGCTCGGAAAGAGAAAGTTCGGCCCTCGAT 492
Db 766 CGGAAAGAGTCTTACAGAGCTCCGCCAGCTCGGAAAGAGAAAGTTCGGCCCTCGAT 825
QY 493 GAGATCATGGAGCTCGAAGAGGAAAGAAAGGACCGCACGACAGAGCCCTGGTTACAG 552
Db 826 GAGATCATGGAGCTCGAAGAGGAAAGAAAGGACCGCACGACAGAGCCCTGGTTACAG 885
QY 553 CCGGGATCGTTGTGAAAAATTTATAACGAAGAAAGCTTGGGGAGAAATATACAAAGAGAAA 612
Db 886 CCGGGATCGTTGTGAAAAATTTATAACGAAGAAAGCTTGGGGAGAAATATACAAAGAGAAA 945
QY 613 GGGGTCGTTAAGGAAGTGAATGACAGAGTACACAGCTGTGGTAAAGATGACTGCTCGA 672
Db 946 GGGGTCGTTAAGGAAGTGAATGACAGAGTACACAGCTGTGGTAAAGATGACTGCTCGA 1005
QY 673 GACAGGCTGAAACTGGACGACACTCATTTAGAGACAGTCATTCGGGCCCGGGGAAAAAGG 732
Db 1006 GACAGGCTGAAACTGGACGACACTCATTTAGAGACAGTCATTCGGGCCCGGGGAAAAAGG 1065
QY 733 GTTCTAGTTTTAAATGGAGGCTACAGAGGAAATGAAGGCACTCTCGAATCCATCAATGAG 792
Db 1066 GTTCTAGTTTTAAATGGAGGCTACAGAGGAAATGAAGGCACTCTCGAATCCATCAATGAG 1125
QY 793 AAGGCTTTTTCAGCCACGATAGTCAATGAAACTGGAACCTTTGAAAGGACGAGAGTTGAA 852
Db 1126 AAGGCTTTTTCAGCCACGATAGTCAATGAAACTGGAACCTTTGAAAGGACGAGAGTTGAA 1185
QY 853 GGTATTCATATGAAGACATATCTAAACTGCTTGAAGTTTGAAGATTTGATAACACACA 912
Db 1186 GGTATTCATATGAAGACATATCTAAACTGCTTGAAGTTTGAAGATTTGATAACACACA 1245
QY 913 TTGAACTGTGAAGCATCAAAATTTGGTGTAGCCAGGCACTGTGTAACTCTACTGTGTTA 972
Db 1246 TTGAACTGTGAAGCATCAAAATTTGGTGTAGCCAGGCACTGTGTAACTCTACTGTGTTA 1305
QY 973 GGGGATTTGTTTGTATTTAAAAAATAATCATCTATTAAATACACTAGTGAATAGTTG 1032
Db 1306 GGGGATTTGTTTGTATTTAAAAAATAATCATCTATTAAATACACTAGTGAATAGTTG 1365
QY 1033 GGTAAATTTATAAATAATCATCTATTAAATACACTAGTGAATAGTTG 1092
Db 1366 GGTAAATTTATAAATAATCATCTATTAAATACACTAGTGAATAGTTG 1425
QY 1093 AAAAAA 1102
Db 1426 AAAAAA 1435
RESULT 4
AX85551
ID AX85551 standard; cDNA; 1002 BP.
XX AX85551;
AC AX85551;
XX 07-SEP-1999 (first entry)
XX cDNA of a gene coding for a truncated human kin17 protein.
DE
XX

kin17 protein; cell proliferation; fertility; hyperproliferative disease;
protein interaction; curved DNA; HIV replication; HIV integration;
repair enzyme; ss.
Homo sapiens.
FR2772046-A1.
11-JUN-1999.
09-DEC-1997; 97PR-00015536.
09-DEC-1997; 97PR-00015536.
(COMS) COMMISSARIAT ENERGIE ATOMIQUE.
Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
WPI; 1999-359999/31.
New DNA coding for human kin17 protein - useful for controlling cell
proliferation or fertility.
Claim 5; Page 32; 69pp; French.
The present sequence encodes a truncated human kin17 protein with amino
acids 129-228 deleted. The mammalian kin17 protein is useful for
preparing a medicament for controlling cell proliferation or for
controlling fertility. The medicaments can also be used to treat
hyperproliferative diseases. Fragments between amino acids 55 and 235
(preferably between amino acids 129 and 228) of a mammalian kin17 protein
are useful for regulating the interaction between proteins and curved
DNA. The fragment can be used to block replication of HIV or its
integration into the human genome or to target repair enzymes to curved
DNA sites. Expression vectors for kin17 can be used for controlling cell
proliferation

Query Match 63.9%; Score 704.6; DB 2; Length 1002;
Best Local Similarity 86.1%; Pred. No. 6.8e-130;
Matches 819; Conservative 0; Mismatches 119; Indels 13; Gaps 3;
1 ATGGGCAAGTCGGATTTCTGAGCCCAAGCCATCCCAATAGATTGACCAAGGG 60
58 ATGGGCAAGTCGGATTTCTTACTCCCAAGCTATGCCCAACAGGATCAAGTCCAAAGGG 117
61 CTCCAGAGCTTCGCTGTACTGCGATGTCGCAAGCAATGCCGCGACGAGAATGGC 120
118 CTGCAAGAGCTTACCGTGTATGTCAGATGTCGCAAGCAATGCCGCGACGAGAATGGC 177
121 TTTAAGTGCTACTGTATGCTGAATCTCATCAAGACAACTGTTGCTGGCTTCAGAAAC 180
178 TTTAAGTGCTACTGTATGCTGAATCTCATCAAGACAACTGTTGCTGGCTTCAGAAAC 237
181 CCTCAGAGCTTATGGAATTTTTCAGAGAAATCCGAATGACTTCTCGAACTTCTG 240
238 CCTCAGAGCTTATGGAATTTTTCAGAGAAATCCGAATGACTTCTCGAACTTCTC 297
241 AGGCGAGCTTGGCACTTAAAGGGTCCACAAACAACTGTTCAATGAATACATCAGC 300
298 AGGAGAGCTTGGCACTTAAAGGGTCCACAAACAACTGTTCAATGAATACATCAGC 357
301 CACGAGAGCACTCCACATGAACGCTACCCAGTGGGAGACACTGACCGACTTTACCAAG 360
358 CACGAGAGCACTCCACATGAATGCCACTCAGTGGGAACTCTGACTGATTTTACTAAG 417
361 TGGCTGGGAGAGAGGGCTTGTAAAGGTACAGTGCAGTGCAGTGCCTGGGAGCGCA 420
418 TGGCTGGGAGAGAGGGCTTGTGCAAA-----AGTGCAGTGAAGACATAGGAAGTTCA 471
421 GCATCCGGGAACGGAAGAGTCTTCAGAGAGCTCCGCCAGCT-----GCCGAAGAG 474

472 GCATCAGTGAACGAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAGAGAAA 531
475 AAGAGCTCGCCCTCGATGAGATCATGGAGCTCGAAGAGAGAAAGAAAGGACCGCACGG 534
532 AAGAAATCTGCACTGGATGAATCATGGAGATTGAAGAGGAAAGAAAGAACTGCCCCGA 591
535 ACAGAGCCCTGGTTACAGCGGGGATCGTTGTGAAAAATTATACGAAGAAAGCTTGGGGAG 594
592 ACAGAGCTACTGGCTACAGCTGAAATTAATTGTGAAAAATTATACGAAGAACTGCGGAGAG 651
595 AATATCAAGAGAAAGGGGCTGTTAAGGAGTCAATGACAGGTACACAGCTGTGGTA 654
652 AATATCAAGAGAAAGGGCTGTTAAGGAGTCAATGACAGGTACACAGCTGTGGTA 711
655 AAGATGACTGACTCTCGAGACAGGCTGAAACTGACAGACTCATTTAGAGACAGTCAAT 714
712 AAGATGATGATCTCGAGACAGGCTGAAACTGACAGACTCATTTAGAGACAGTCAAT 771
715 CCGGCCCCGGGAAAGGGTCTAGTTTAAATGGAGGCTACAGAGAAATGAAGGCACT 774
772 CCAGCACCAGGAAAGAAATTTCTAGTTTAAATGGAGGCTACAGAGAAATGAAGGTACC 831
775 CTCGAATCCATCAATGAGAGGCTTTTTCAGCCAGCATAGTCAATGAAACTGGACCTTTG 834
832 CTAGAATCCATCAATGAGAGGCTTTTTCAGCTACTATCGTCAATGAAACTGGCCCTTA 891
835 AAGAGGACGAGAGTTCGAAGTATTCAATATGAAGACATATCTAAACTTGTCTGAGTTGA 894
892 AAGAGGACGAGAGTTCGAAGTATTCAATATGAAGACATTTCTAAACTTGTCTGAGTTGA 951
895 AATTTGTTAATCAACATCACTTTAAATCTTAAAGCATCAAAATTTGTTGTCG 944
952 AATTTGTTAATCAACATCACTTTAAATCTTAAAGCATCAAAATTTGTTGTCG 1002

RESULT 5
AA85549
ID AA85549 standard; cDNA; 1296 BP.
AC AA85549;
XX 07-SEP-1999 (first entry)
DT
XX cDNA of a gene coding for the human kin17 protein.
DE
XX Human; kin17 protein; cell proliferation; fertility;
KW hyperproliferative disease; protein interaction; curved DNA;
KW HIV replication; HIV integration; repair enzyme; ss.
XX Homo sapiens.
OS
XX FR2772046-A1.
PN
XX 11-JUN-1999.
PD
XX 09-DEC-1997; 97PR-00015536.
PF
XX 09-DEC-1997; 97PR-00015536.
PR
XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
PA
XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
PI
XX WPI; 1999-359999/31.
XX
DR New DNA coding for human kin17 protein - useful for controlling cell
XX proliferation or fertility.
XX
PT Claim 1; Page 30-31; 69pp; French.
XX
PS The present sequence encodes a human kin17 protein. The mammalian kin17
XX protein is useful for preparing a medicament for controlling cell
CC proliferation or for controlling fertility. The medicaments can also be
CC

Db 827 AAACCTTGACCACTCATTTAGAGCAGTAATTCAGACACGAGAAAGAAATCTAGTT 886
QY 742 TTAATGAGGCTACAGAGAAATGAAGCACTCTCGAATCCATCAATGAGAAGGCTTTT 801
Db 887 TTAATGAGGCTACAGAGAAATGAAGTACCTAGATCCATCAATGAGAAGACTTTT 946
QY 802 TCAGCCACATAGTCATTCGAACCTGACCTTTGAAGGACGAGATTTGAAGGTAT 857
Db 947 TCAGCTACTATCGTCATTCGAACTGGCCCTTTAAAGAGGACGAGAGTTGAAGGAAT 1002

RESULT 8
AAH34095
ID AAH34095 standard; cDNA; 591 BP.
XX
AC AAH34095;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:1177.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; ss.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US026524.
XX
PR 29-SEP-1999; 99US-0157137P.
PR 03-NOV-1999; 99US-0163280P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
XX WPI: 2001-235357/24.
DR P-PSDB; AAG74690.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers.
XX
PS Claim 1; Page 2977-2978; 9803pp; English.

CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene therapy
CC and vaccine production. N and P may be used in the prevention, diagnosis
CC and treatment of diseases associated with inappropriate P expression. For
CC example, N and P may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of P by expressing inactive proteins or to
CC supplement the patients own production of P. Additionally, N may be used
CC to produce the colon cancer-associated Ps, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the proteins. N and P
CC can be used in the prevention, diagnosis and treatment of colorectal
CC carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent
CC sequences used in the exemplification of the present invention. N.B.
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC time of publication, meaning no sequences are present for SEQ ID NO:1027
CC to 1052, 7921 and 7922

XX Sequence 591 BP; 199 A; 95 C; 114 G; 182 T; 0 U; 1 Other;

Query Match 24.4%; Score 269.2; DB 4; Length 591;
Best Local Similarity 77.8%; Pred. No. 9.4e-44;
Matches 325; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 646 GCTGTGTAAGATGACTGACTCTCGAGACAGGCTGAACTGGACACAGCTCATTTAGAG 705
Db 28 GCTGAGGAATTCGGCACGAGTCTGGAGACAAGCTGAAACTTGACCAGACTCATTTAGAG 87
QY 706 ACAGTCATTCCGGCCCCGGGAAAGGGTTCTAGTTTTTAAATGGAGGCTACAGAGGAAAT 765
Db 88 ACAGTAATTCAGCACACCAGGAAAGAAATTCAGTTTTTAAATGGAGGCTACAGAGGAAAT 147
QY 766 GAAGCAGCTCTCGAATCCATCAATGAGAAAGGCTTTTTCAGCCACGATAGTCATTTGAAACT 825
Db 148 GAAGGTACCTAGATCCATCAATGAGAAGACTTTTTCAGCTACTATCGCTCATTTGAAACT 207
QY 826 GGACCTTTGAAGGACGAGAGTTGAAGGTATTCATATGAAGACATATCTAAACTTGTCT 885
Db 208 GGCCCTTTAAAGGACGAGAGTTGAAGGAATTCATATGAAGACATTTCTAAACTTGTCT 267
QY 886 TGAGTTTGAATAATTTGATAACAACACATTGAAACTGTGAAAGCATCAAAATTTGGTGTAGCC 945
Db 268 TGAGTTTGAATAATTTGTTAAACAATACATTAATACTTAAAGCATCAAAATTTGGTGTAGCC 327
QY 946 AAGGCAGCTGTGTAAGTCTACTGTGTAGGGGATTTTGTATTAATAAAAAAATC 1005
Db 328 AAGGCATTTATGAGACTCTACTGTGTAGGGTATATTTCTTTGTATAAAACAAACAGGTTT 387
QY 1006 ATCTATTTAAATACTAGTGAATAGTTGGGTAAATTTATAATAAATCTATGTTTTTTT 1063
Db 388 TTGAAATAATTTACTGTATAGTTAGTTGTTCAGCTAAACTTTTGAGAAGAAATTTAATAT 445

RESULT 9
ADP22451
ID ADP22451 standard; DNA; 1372 BP.
XX
AC ADP22451;
XX
DT 12-AUG-2004 (first entry)
XX
DE Sea-squirt (Clona intestinalis) zinc finger protein coding sequence #13.
XX
KW sea-squirt; zinc finger protein; gene detection; drug development;
KW zinc finger protein-associated disease; gene; ds.
XX
OS Clona intestinalis.
XX
PN JP2004057126-A.
XX
PD 26-FEB-2004.
XX
PF 31-JUL-2002; 2002JP-00222484.
XX
PR 31-JUL-2002; 2002JP-00222484.
XX
PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX
DR WPI: 2004-208711/20.
DR P-PSDB; ADP22452.
XX
PT Novel gene encoding zinc finger protein, useful as probe in gene
PT detecting instruments and in development of drug for treating zinc finger
PT protein associated diseases.
XX
PS Claim 3; SEQ ID NO 25; 972pp; Japanese.
XX
CC The invention comprises the amino acid and coding sequences of sea-squirt
CC (Clona intestinalis) zinc finger proteins. The DNA and protein sequences
CC of the invention are useful in a gene detecting instrument. The DNA and
CC protein sequences of the invention are useful in the development of drugs
CC for the treatment of zinc finger protein-associated diseases. The present
CC DNA sequence encodes a sea-squirt zinc finger protein of the invention.
SQ Sequence 1372 BP; 500 A; 224 C; 286 G; 362 T; 0 U; 0 Other;
Query Match 17.2%; Score 190; DB 12; Length 1372;

Best Local Similarity 68.6%; Pred. No. 5.1e-28;		Matches 262; Conservative 0; Mismatches 120; Indels 0; Gaps 0;	
QY	7	AAGTCGGATTTCTGTAGCCCAAGCCATGCGCAATAGAAATTAAGTCCAAAGGGCTCCAG	66
Db	55	AAACCAAGGATTTCTCACTGCAAGGCTATCGGCACCGTATCAATCAAAAGGGACTTCAA	114
QY	67	AAGCTTCGGTGTGCTACTGCGCAGATGTGCCAAAGCAATGCCGACGAGATGCGCTTTAAG	126
Db	115	AAGTTCGGATGGTATTGTCAATGTGCCAAACACGTGCGAGATGAGAACGGGTTTAA	174
QY	127	TGTCACGTGTATGCTGAATCTCATCAAGACCACTGTGTGCTTCAGAAACCCCTCAG	186
Db	175	TGCCATGTATGTCGCAATCACATAGCGCCAAATGTTAATGCTGCTGAAATCCTGGA	234
QY	187	CAGTTTATGGATTAATTTTCAGAGGAATCCGAAATGACTTTCTGGAACCTTCAGAGCGGA	246
Db	235	GCCTTTTCACAGTTCTCTTTCAAGTCTCTTTTCAAGAGCTTTATGCAACTTTTGAAACCA	294
QY	247	CGCTTTGGCACTAAAGGGTCCACACACATTTGTCTACAAATGAATACATCAGCCACCGA	306
Db	295	CGTTTCGGGACTAAAGAGTTCAATCAATATTTGTGTATACGAGTATATTTCCCATAA	354
QY	307	GAGCACATCCATGAACGCTACCCAGTGGGAGACACTGACCGACTTTACCAAGTGGCTG	366
Db	355	GAACACATTCATATGACGCTCAAAATGGGTGACGTTAACTGCTTACAAATGCTTG	414
QY	367	GGCAGAGAGGGCTTGTGTAAG	388
Db	415	GGTAGAGAGGGGACACTGTAAAG	436
RESULT 10			
AAC47248			
ID	AAC47248 standard; DNA; 1459 BP.		
XX	AAC47248;		
AC	AAC47248;		
XX	18-OCT-2000 (first entry)		
DT	Arabidopsis thaliana DNA fragment SEQ ID NO: 53113.		
DE	Arabidopsis thaliana		
XX	Hybridisation assay; genetic mapping; gene expression control;		
KW	protein identification; signal transduction pathway; metabolic pathway;		
KW	promoter; termination sequence; ss.		
XX	Arabidopsis thaliana		
OS	Arabidopsis thaliana		
XX	EP1033405-A2.		
PN	06-SEP-2000.		
PD	25-FEB-2000; 2000EP-00301439.		
XX	25-FEB-1999; 99US-0121825P.		
XX	05-MAR-1999; 99US-0123180P.		
PR	09-MAR-1999; 99US-0123548P.		
PR	23-MAR-1999; 99US-0125788P.		
PR	25-MAR-1999; 99US-0126264P.		
PR	29-MAR-1999; 99US-0126785P.		
PR	01-APR-1999; 99US-0127462P.		
PR	06-APR-1999; 99US-0128234P.		
PR	08-APR-1999; 99US-0128714P.		
PR	16-APR-1999; 99US-0129845P.		
PR	19-APR-1999; 99US-0130077P.		
PR	21-APR-1999; 99US-0130449P.		
PR	23-APR-1999; 99US-0130510P.		
PR	28-APR-1999; 99US-0130891P.		
PR	30-APR-1999; 99US-0131449P.		
PR	30-APR-1999; 99US-0132048P.		
PR	04-MAY-1999; 99US-0132407P.		
PR	05-MAY-1999; 99US-0132484P.		
PR	05-MAY-1999; 99US-0132485P.		

PR	06-MAY-1999;	99US-0132486P.
PR	06-MAY-1999;	99US-0132487P.
PR	07-MAY-1999;	99US-0132863P.
PR	11-MAY-1999;	99US-0134256P.
PR	14-MAY-1999;	99US-0134218P.
PR	14-MAY-1999;	99US-0134219P.
PR	14-MAY-1999;	99US-0134221P.
PR	14-MAY-1999;	99US-0134370P.
PR	18-MAY-1999;	99US-0134768P.
PR	19-MAY-1999;	99US-0134941P.
PR	20-MAY-1999;	99US-0135124P.
PR	21-MAY-1999;	99US-0135353P.
PR	24-MAY-1999;	99US-0135629P.
PR	25-MAY-1999;	99US-0136021P.
PR	27-MAY-1999;	99US-0136392P.
PR	28-MAY-1999;	99US-0136782P.
PR	01-JUN-1999;	99US-0137222P.
PR	03-JUN-1999;	99US-0137528P.
PR	04-JUN-1999;	99US-0137502P.
PR	07-JUN-1999;	99US-0137724P.
PR	08-JUN-1999;	99US-0138094P.
PR	10-JUN-1999;	99US-0138540P.
PR	10-JUN-1999;	99US-0138847P.
PR	14-JUN-1999;	99US-0139119P.
PR	16-JUN-1999;	99US-0139452P.
PR	17-JUN-1999;	99US-0139453P.
PR	17-JUN-1999;	99US-0139492P.
PR	18-JUN-1999;	99US-0139454P.
PR	18-JUN-1999;	99US-0139455P.
PR	18-JUN-1999;	99US-0139456P.
PR	18-JUN-1999;	99US-0139457P.
PR	18-JUN-1999;	99US-0139458P.
PR	18-JUN-1999;	99US-0139459P.
PR	18-JUN-1999;	99US-0139460P.
PR	18-JUN-1999;	99US-0139461P.
PR	18-JUN-1999;	99US-0139462P.
PR	18-JUN-1999;	99US-0139463P.
PR	18-JUN-1999;	99US-0139750P.
PR	18-JUN-1999;	99US-0139763P.
PR	21-JUN-1999;	99US-0139817P.
PR	23-JUN-1999;	99US-0139899P.
PR	23-JUN-1999;	99US-0140353P.
PR	23-JUN-1999;	99US-0140354P.
PR	24-JUN-1999;	99US-0140695P.
PR	28-JUN-1999;	99US-0140823P.
PR	29-JUN-1999;	99US-0140921P.
PR	30-JUN-1999;	99US-0141287P.
PR	01-JUL-1999;	99US-0141842P.
PR	02-JUL-1999;	99US-0142055P.
PR	06-JUL-1999;	99US-0142390P.
PR	08-JUL-1999;	99US-0142803P.
PR	09-JUL-1999;	99US-0142920P.
PR	12-JUL-1999;	99US-0142977P.
PR	13-JUL-1999;	99US-0143542P.
PR	14-JUL-1999;	99US-0143624P.
PR	15-JUL-1999;	99US-0144005P.
PR	16-JUL-1999;	99US-0144085P.
PR	19-JUL-1999;	99US-0144086P.
PR	19-JUL-1999;	99US-0144325P.
PR	19-JUL-1999;	99US-0144331P.
PR	19-JUL-1999;	99US-0144332P.
PR	19-JUL-1999;	99US-0144333P.
PR	19-JUL-1999;	99US-0144334P.
PR	19-JUL-1999;	99US-0144335P.
PR	20-JUL-1999;	99US-0144352P.
PR	20-JUL-1999;	99US-0144632P.
PR	20-JUL-1999;	99US-0144884P.
PR	21-JUL-1999;	99US-0144814P.
PR	21-JUL-1999;	99US-0145086P.
PR	21-JUL-1999;	99US-0145088P.
PR	22-JUL-1999;	99US-0145085P.
PR	22-JUL-1999;	99US-0145087P.

PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 10-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154079P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159684P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.

PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 15.8%; Score 174; DB 3; Length 1459;
Best Local Similarity 65.4%; Pred. No. 7.5e-25;
Matches 255; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

Qy 1 ATGGGCAAGTCGGATTTTCTGAGCCCAAGGCCATCGCCAATAGAAATTAAGTCCAAAGG 60
Db 1 ATGGGTAAGAACGATTTTCTAAACCCCAAGGCGATTGCGAATCGAATTAAGCCCAAGGA 60

Qy 61 CTCGAGAGCTTCGCTGCTACTGCCAGATGTGCCAAAGCAATGCCGACGAGAAATGCG 120
Db 61 CTTCAAAAGCTTCGATGGTATTGTTCAGATGTTCAGAAACAATGCCGAGACGAGAAATGGA 120

Qy 121 TTTAAGTGTCACTGTATGTCTGAATCTCATCAAAGACAATCTGTTGCTGGCTTCAGAAAC 180
Db 121 TTCAGTGTCAATTGTATGAGCGAATCTCACCAGAGGCGAGATGCAAGTTTTCGGGCGAAT 180

Qy 181 CCTCAGCAGTTTATGGATTATTTTCAGAGAAATTCGAAATGATCTTCTGGAATTCGTG 240
Db 181 CCGACTCGTTGTTCGATGGTTACTCTGAGAAATTCGAGCAGACGCTTCTCGATCTGATG 240

Qy 241 AGGCGACGCTTGGCACTAAAGGGTCCACAACAACATTTGTACAATGAATACATCAGC 300
Db 241 AGGCGGAGCCACCGTTTCTCGTATCGCTGCCACTGTGTCACATGAGTATATTAC 300

Qy 301 CACCGAGAGCACATCCACATGAACGCTACCCAGTGGGAGACACTGACCGACTTTTACCAG 360
Db 301 GATAGACATCATGTTTCATCATGAATTCGACGGAGTGGGCGGCGAGTTTATCAAG 360

Qy 361 TGGCTGGCGAGAGAGGGCTTGTGTAAGGT 390
Db 361 CATCTTGGGAAAACCTGTAAGTTAAGGTT 390

RESULT 11
AD084255
ID AD084255 standard; cDNA; 1614 BP.
XX AC AD084255;
XX DT 21-APR-2005 (first entry)
XX DE Plant full length insert polynucleotide seqid 2975.
XX KW plant protectant; plant growth regulant; gene therapy; plant;
KW recombinant DNA construct; physical array; plant breeding marker;
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW extreme osmotic condition; pathogen tolerance; pest tolerance;
KW growth rate; cell cycle pathway; disease resistance;
KW galactomannan production; lignin production; plant growth regulator;
KW yield; plant growth; plant development; seed oil; protein yield;
KW protein content; gene; ss.
XX OS Unidentified.
XX US US2004034888-A1.
XX PN 19-FEB-2004.

```
XX 28-APR-2003; 2003US-00425114.
PF
XX
PR 06-MAY-1999; 99US-00304517.
PR 05-NOV-2001; 2001US-00985678.
XX
PA (LIU/) LIU J.
PA (ZHOU/) ZHOU Y.
PA (KOVA/) KOVALIC D K.
PA (SCRE/) SCREEN S E.
PA (TABA/) TABASKA J E.
PA (CAO/) CAO Y.
XX
XX Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
XX WPI; 2004-180133/17.
XX
XX New recombinant DNA construct, useful for improving plant tolerance to
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT pests, for conferring increased resistance to plant disease, or for
PT improving yield.
XX
XX Claim 1; SEQ ID NO 2975; 15pp; English.
XX
XX The invention describes a recombinant DNA construct comprising a
CC polynucleotide consisting of a sequence encoding an amino acid sequence
CC available in electronic form from the US patent office at
CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC increased resistance to plant disease, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This sequence represents a plant full length insert
CC polynucleotide that can be used in the recombinant DNA construct of the
CC invention.
XX
XX Sequence 1614 BP; 399 A; 362 C; 493 G; 360 T; 0 U; 0 Other;
SQ
Query Match 15.3%; Score 168.8; DB 13; Length 1614;
Best Local Similarity 64.7%; Pred. No. 8.2e-24;
Matches 251; Conservative 0; Mismatches 137; Indels 0; Gaps 0;
QY 1 ATGGGCAAGTCGGATTTCTGAGCCCAAGCCCATGCCCAATAGATTAAATGTCGAAGGG 60
DB 80 ATGGGGAAGCAGCAGTTCCTGACCCGGAAGCGCATCGCAACCGGATCAAGGGCAAGGG 139
QY 61 CTCAGAGCTTCGCTCGTACTGTCAGATGTCGCAAAAGCAATGCCCGCAGCAATGGC 120
DB 140 CTGAGAGCTTCGCGTGCTGTCAGATGTCGAGAGCGATGCCCGCAGCAAGCGG 199
QY 121 TTTAAGTGTCATCTGATGTCGTAATCTCATCAAGACAACCTGTTGCTGGCTTCAGAAAC 180
DB 200 TTCAAGTGCCATCGATGTCGAGTCGACCCAGAGGCGAGTGCAGGTGTTCCGCGATGCC 259
QY 181 CCTCAGCAGTTATGGAATATTTTTCAGAGAAATTCGAAATGATCTTCTGGAACCTCTG 240
DB 260 CCCGACCGGTCGTCGAGGGCTTCTCCGAGGAGTTCCTCGAGTCTCTTCTCTCCCTCATC 319
QY 241 AGGCGACGCTTGGCACTAAAAGGTCACAAACACATTTGCTCAATGAATACATCAGC 300
DB 320 CGCCGCGCGCACCGCCACTCCCGCGTCGCCGCCACCGTCGCTTACAACGAGTACATGCC 379
QY 301 CACCGAGAGCATCTCCACATGAAGCGTACCAGTGGGAGACACTGACCGACTTTTACCAAG 360
DB 380 GACCGGACCACTCCACATGAATCCACGGGCTGGGGCCAGCTCACCAGTTCGTCAAG 439
```

361 TGGCTGGCAGAGAGGGCTTGTGTAAG 388
440 TTCCTGGGCGGAGGGGTACTGTGAAG 467

RESULT 12
AD47874
ID AD47874 standard; cDNA; 1713 BP.
XX
AC AD47874;
XX
DT 21-APR-2005 (first entry)
XX
XX Plant full length insert polynucleotide seqid 22614.
XX
XX plant protectant; plant growth regulant; gene therapy; plant;
KW recombinant DNA construct; physical array; plant breeding marker;
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW extreme osmotic condition; pathogen tolerance; pest tolerance;
KW growth rate; cell cycle pathway; disease resistance;
KW galactomannan production; lignin production; plant growth regulator;
KW yield; plant growth; plant development; seed oil; protein yield;
KW protein content; gene; ss.
XX
XX Unidentified.
XX
XX US2004034888-A1.
XX
XX 19-FEB-2004.
XX
XX 28-APR-2003; 2003US-00425114.
XX
XX 06-MAY-1999; 99US-00304517.
XX 05-NOV-2001; 2001US-00985678.
XX
XX (LIU/) LIU J.
XX (ZHOU/) ZHOU Y.
XX (KOVA/) KOVALIC D K.
XX (SCRE/) SCREEN S E.
XX (TABA/) TABASKA J E.
XX (CAO/) CAO Y.
XX
XX Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
XX WPI; 2004-180133/17.
XX
XX New recombinant DNA construct, useful for improving plant tolerance to
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT pests, for conferring increased resistance to plant disease, or for
PT improving yield.
XX
XX Claim 1; SEQ ID NO 2975; 15pp; English.
XX
XX The invention describes a recombinant DNA construct comprising a
CC polynucleotide consisting of a sequence encoding an amino acid sequence
CC available in electronic form from the US patent office at
CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC increased resistance to plant disease, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This sequence represents a plant full length insert
CC polynucleotide that can be used in the recombinant DNA construct of the
CC invention.
XX
XX Sequence 1713 BP; 423 A; 422 C; 504 G; 364 T; 0 U; 0 Other;
SQ

The invention describes a recombinant DNA construct comprising a polynucleotide consisting of a sequence encoding an amino acid sequence available in electronic form from the US patent office at <ftp.segdata.uspto.gov/sequence.html?DocID:2004034888>. The polynucleotide of the invention are also useful in physical arrays of molecules and as plant breeding markers. The recombinant DNA construct is useful for improving plant tolerance to cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or pests, for manipulating growth rate in plant cells by modification of the cell cycle pathway, for conferring increased resistance to plant disease, for increasing the rate of homologous recombination in plants, for improving yield by modification of photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake or by providing improved plant growth and development under at least one stress condition or for modifying seed oil or protein yield and/or content. This sequence represents a plant full length insert polynucleotide that can be used in the recombinant DNA construct of the invention.

Query Match	15.3%;	Score 168.8;	DB 13;	Length 1713;
Best Local Similarity	64.7%;	Pred. No. 8.3e-24;		
Matches 251;	Conservative 0;	Mismatches 137;	Indels 0;	Gaps 0;
QY	1	ATGGGCAAGTCGGATTCTCTGAGCCCAAGGCCATCGCCCAATAGATTAAAGTCCAAAGGG	60	
DB	239	ATGGGGAACACAGATTCTCTGACGCCGAAGGCGATCGCAACCGGATCAAGCGGAAGGGG	298	
QY	61	CTCCGAAGCTTCGCTGGTACTGCCAGATGTGCCAAAGCAATGCCGCGACGAGAATGGC	120	
DB	299	CTGCAGAAGCTCGGTGGTACTGTCTAGATGTGTCTAGAACGAGTGGCGGACGAGAACGGG	358	
QY	121	TTTAAGTGTCACTGTATGTCTGAATCTCATCAAAAGACAACCTGTGTCTGGCTTCAGAAAC	180	
DB	359	TTCAAAGTGCACATGTCTGAGGTCTCGAGTGCACACGAGGCGAGATGTTCTCGCATGGCC	418	
QY	181	CCTCAGCAGTTCATGGATTATTTTTCAGAGGAATTCGGAATGACCTTTCTGGAATCTCG	240	
DB	419	CCCGACCGGTCTGAGGGCTTCTCCGAGGAGTTCCTCGAGTCCCTCTCCCTCATC	478	
QY	241	AGCGCAGCTTTGGCACTAAAGGGTCCCAACAACATTTGTCTACAATGAATACATCAGC	300	
DB	479	CGCGCGGCACCGCCACTCCCGCTGCGCGCCACCGTCTCTACAACGAGTACATCGCC	538	
QY	301	CACCGAGACGATCCACATGAACGCTACCCAGTGGGAGACACTGACCGACTTTTACCAG	360	
DB	539	GACCGGCACCATCTCCACATGAACCTCCACGCGGTGGGCCACGCTCACCAGTTCGTCAAG	598	
QY	361	TGGCTGGGCAGAGAGGGCTTGTAAAG	388	
DB	599	TTCCCTGGGCGCGAGGGGTACTGTAAGG	626	
RESULT 13				
AA52589				
ID	AA52589	standard; cDNA; 1764 BP.		
XX	AA52589;			
AC				
XX				
DT	27-SEP-2000	(first entry)		
XX				
DE	cDNA encoding maize KIN17 orthologue, ZmKINH-1.			
XX				
KW	ZmKINH-1; KIN17 orthologue; maize; zinc finger protein; RecA homologue;			
KW	nuclear localisation; nonhomologous recombination;			
KW	illegitimate recombination; double stranded DNA binding; curved DNA;			
XX	homologous gene targeting; transgenic plant; ss.			
OS	Zea mays.			
XX				
FH	Key	Location/Qualifiers		
FT	CDS	208..1482		
PT		/tag= a		
FT		/product= "Maize KIN17 orthologue, ZmKINH-1"		
FT		/function= "Binds double-stranded curved DNA"		
XX				
PN	WO20024900-A1.			
PD	04-MAY-2000.			
XX				
PF	06-OCT-1999;	99WO-US023280.		
XX				
PR	27-OCT-1998;	98US-0105802P.		
XX				
PA	(PION-) PIONEER HI-BRED INT INC.			
XX				
PI	Mahajan PB;			
XX				
DR	WPI; 2000-350741/30.			
DR	P-PSDB; AAB03063.			
XX				
PT	Nucleic acids encoding maize KIN17 orthologue proteins useful for			

preventing illegitimate recombination in cells.

Claim 1; Page 71-73; 84pp; English.

This sequence represents cDNA encoding the maize KIN17 orthologue ZmKINH-1. The invention relates to maize KIN17 orthologues (AAB03063- AAB03065), nucleic acids encoding them (AA52589-A52591), and expression vectors, transgenic plants and plant seeds comprising nucleotides encoding maize KIN17 orthologues. KIN17 has, until now, been found only in animal (avian, rodent and human) cells, this invention being the first report describing the presence of KIN17 in plants. Murine KIN17 was found to have significant homology to Escherichia coli RecA protein, and contains a zinc finger motif and a nuclear localisation signal. KIN17 binds double stranded DNA, preferentially binding to curved DNA, and forms intranuclear foci on overexpression in mammalian cells. It is also induced on exposure to gamma or ultraviolet radiation. These findings indicate that KIN17 plays a role in non-homologous (illegitimate) recombination, which occurs at higher rates among higher eukaryotes, particularly plants. Illegitimate recombination in plants is a major impediment to the generation of transgenic crops such as maize. Maize KIN17 orthologue nucleic acid sequences may be used to generate transgenic plants. The transgenic plants generated can be monocots or dicots and are particularly maize, soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton, rice, barley and millet. In particular, antisense sequences may be used to reduce KIN17 levels in embryogenic callus or embryo cells, thereby reducing the amount of non-homologous recombination and enhancing homologous gene targeting

Sequence 1764 BP; 434 A; 427 C; 516 G; 387 T; 0 U; 0 Other;

Query Match 15.3%; Score 168.8; DB 3; Length 1764;

Best Local Similarity 64.7%; Pred. No. 8.3e-24;

Matches 251; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 1 ATGGGCAAGTCGGATTCTCTGAGCCCAAGGCCATCGCCCAATAGATTAAAGTCCAAAGGG 60

DB 208 ATGGGGAACACAGATTCTCTGACGCCGAAGGCGATCGCAACCGGATCAAGCGGAAGGG 267

QY 61 CTCCGAAGCTTCGCTGGTACTGTCCAGATGTGCCAAAGCAATGCCGCGACGAGAATGGC 120

DB 268 CTGCAGAAGCTCGGTGGTACTGTCTAGATGTGTCTAGAAAGCAGTGCCTCGGACGAGAACGG 327

QY 121 TTTAAGTGTCACTGTATGTCTGAACTCTCATCAAGACAACCTGTTGCTGCTTCAGAAAC 180

DB 328 TTCAGTCCCACTGCATGTCCGAGTTCGACACGAGGCGAGATGCGGATGGCC 387

QY 181 CCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTCGAAATGACTTTCTGGAACTTCTG 240

DB 388 CCCGACCGGTCGTCGAGGGCTTCTCCGAGGAGTTCCTCGAGTCTCTCTCCCTCATC 447

QY 241 AGGCGACGCTTTGGCACTAAAGGGTCCCAACAAATGTTCTACAATGAAATACATCAGC 300

DB 448 CGCCGCGGCCACCGCCACTCCCGCGTCCGCGCCACCGTCTCTACACGAGTACATCGCC 507

QY 301 CACCGAGAGCATCCACATGACGCTACCCAGTGGGAGACACTGACCGACTTTTACCAG 360

DB 508 GACCGGCACACGCTCCACATGAACCTCCACGCGGTGGGCCACGCTCACCAGTTCGTCAAG 567

QY 361 TGGCTGGGCAGAGAGGGCTTGTAAAG 388

DB 568 TTCCTGGGCGCGAGGGGTACTGTAAGG 595

RESULT 14

ADT17629

ID ADT17629 standard; cDNA; 1713 BP.

XX ADT17629;

XX 13-JAN-2005 (first entry)

XX Plant cDNA, Seq ID 2955.

XX

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 12:58:53 ; Search time 6160.01 Seconds
(without alignments)
8370.011 Million cell updates/sec

Title: US-09-555-529-2
Perfect score: 1102
Sequence: 1 agggcagctcggtttctt.....aaaaaaaaaaaaaaaaaaaa 1102

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*
1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_hic.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_est7.*
9: gb_gss1.*
10: gb_gss2.*
11: gb_gss3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	625.6	56.8	677	2	BF321724
C 2	576	52.3	1081	2	BE283156
C 3	559	50.7	560	2	BG072126
C 4	554.2	50.3	739	5	BQ782294
C 5	545.6	49.5	754	5	BQ780218
C 6	522	47.4	601	3	BP774644
C 7	512	46.5	512	3	BM228037
C 8	512	46.5	512	3	BM229452
C 9	496.2	45.0	1154	10	AY412520
C 10	474	43.0	1396	4	CS959808
C 11	462	41.9	513	2	BE225867
C 12	454	41.2	466	2	BS448684
C 13	449	40.7	1543	4	AY609991
C 14	448.2	40.7	457	1	AW045702
C 15	433.4	39.3	447	2	BB749127
C 16	433.4	39.3	864	1	AM037867
C 17	431	39.1	723	3	BQ186497
C 18	429.8	39.0	871	1	AJ819723
C 19	426.6	38.7	526	2	BE114414
C 20	424.4	38.5	430	6	CF617307
C 21	423.8	38.5	800	1	AM038283
C 22	414	37.6	437	2	BB749400

23	413	37.5	682	1	AV721396
24	406.4	36.9	679	1	AM028238
C 25	404	36.7	733	5	EX104466
C 26	402.8	36.6	786	1	AJ819816
C 27	401	36.4	703	1	AM035210
C 28	400	36.3	493	2	BE120977
C 29	398.4	36.2	691	1	AM028686
C 30	397.4	36.1	848	1	AL558810
C 31	397	36.0	754	8	DN755964
C 32	396	35.9	543	6	CB158644
C 33	395.4	35.9	399	5	BY395138
C 34	392.6	35.6	720	7	CK301009
C 35	389.8	35.4	820	6	CF409895
C 36	388	35.2	638	4	AK009429
C 37	386.4	35.1	625	7	CF914851
C 38	384	34.8	501	6	CA560899
C 39	384	34.8	538	6	CA559106
C 40	383.4	34.8	515	1	AA638405
C 41	383.2	34.8	583	2	BF321725
C 42	383.2	34.8	663	1	AJ660240
C 43	383	34.8	698	1	AI650375
C 44	382.4	34.7	1077	5	BUS11663
C 45	379	34.4	391	1	AU041571

ALIGNMENTS

RESULT 1
BF321724/c
LOCUS
DEFINITION
uz64c04.x1 NCI CGAP Mam6 Mus musculus cDNA clone IMAGE:3673830 3', similar to TR:060870 O60870 KIN17 PROTEIN. ; mRNA sequence.
ACCESSION
BF321724
VERSION
BF321724.1
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidae; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 677)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/iresources.shtml
MGI:1434598
High quality sequence stop: 396.
Location/Qualifiers
1. 677
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3673830"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI-CGAP_Mam6"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigator

ORIGIN	providing samples: Jeffrey Green, M.D., NIH									
Query Match	56.8%;	Score	625.6;	DB 2;	Length	677;				
Best Local Similarity	96.3%;	Pred. No.	3.1e-124;							
Matches	653;	Conservative	0;	Mismatches	19;	Indels	6;	Gaps	1;	
QY	388	GGTACAGTGCAC	CTGAAGTGTCTGGGAGCGGACGATCCGGGAAACGGAAGAGTCTTCA	447						
Db	677	GGCCCAAGTGCAC	TGAAGTGTCTGGGAGCGGACGATCCGGGAAACGGAAGAGTCTTCA	618						
QY	448	CAGAGCTCCGCCAGCT	CGGAGAGAGAGTCCGCCCTGGATCAGATCATGAGGCTC	507						
Db	617	CAGAGCTCCGCCAGCT	CGGAGAGAGAGTCCGCCCTGGATCAGATCATGAGGCTC	558						
QY	508	GAAGAGGAAAGAAAGGAC	CGGACGACGACGCTGGTTACAGCCGGGGATCGTTGTG	567						
Db	557	GAAGAGGAAAGAAAGGAC	CGGACGACGACGCTGGTTACAGCCGGGGATCGTTGTG	498						
QY	568	AAATATTAAACGAAGAAGCT	TGGGAGAAATATCAAGAAGAAAGGGTCTGTTAAGGAA	627						
Db	497	ACAATATTAAACGAAGAAGCT	TGGGAGAAATATCAAGAAGAAAGGGTCTGTTAAGGAA	438						
QY	628	GTGATTGACAGGTACACAGCT	GTGTAAAGATGACTGCTCTGGACACAGGCTGAAACTG	687						
Db	437	GTGATTGACAGGTACACAGCT	GTGTAAAGATGACTGCTCTGGACACAGGCTGAAACTG	378						
QY	688	GACCAGACTCATTTAGACAGT	CACTCCGCCCGGGGAAAGGGTCTAGTTTAAAT	747						
Db	377	GACCAGACTCATTTAGACAGT	CACTCCGCCCGGGGAAAGGGTCTAGTTTAAAT	318						
QY	748	GGAGGCTACAGAGGAAATGAAGGCA	CTCTCGAATCCATCAATGAGAAGGCTTTTCAGCC	807						
Db	317	GGAGGCTACAGAGGAAATGAAGGCA	CTCTCGAATCCATCAATGAGAAGGCTTTTCAGCC	258						
QY	808	ACGATAGTCATTGAACCTGGACCT	TTGAAGGACGAGGTTCAAGGTTCAATATGAA	867						
Db	257	ACGATAGTCATTGAACCTGGACCT	TTGAAGGACGAGGTTCAAGGTTCAATATGAA	198						
QY	868	GACATATCTAAACTCTGCTGAGT	TTTGAAATTTGATAACACACATTTGAACTGTGAAGC	927						
Db	197	GACATATCTAAACTCTGCTGAGT	TTTGAAATTTGATAACACACATTTGAACTGTGAAGC	138						
QY	928	ATCAAAATTCGTGTAGCCAGGCACT	GTGTAACTTACTGTGTAGGGAATTTGTTTGT	987						
Db	137	ATCAAAATTCGTGTAGCCAGGCACT	GTGTAACTTACTGTGTAGGGAATTTGTTTGT	78						
QY	988	ATTAAAAAATAAATAATCATCT	ATTTAAATCTAGTGAATAGTTCGGTAAATTTAATA	1047						
Db	77	AT-----	AAAAAATAATCATCTATTTAAATAGTGAATAGTTCGGTAAATTTAATA	24						
QY	1048	AAATCTATGTTTTTTTA	1065							
Db	23	AAATCTATGTTTTTTTAA	6							
RESULT 2	BE283156	601103709F1 NCI_CGAP_Lu29 Mus musculus cdna clone IMAGE:3500835 5', linear EST 26-OCT-2000								
LOCUS	DEFINITION	mRNA sequence.								
ACCESSION	BE283156									
VERSION	BE283156.1	GI:9159216								
KEYWORDS	EST.									
SOURCE	Mus musculus (house mouse)									
ORGANISM	Mus musculus									
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;									
	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;									
	Sciurognathi; Murcidae; Muridae; Murinae; Mus.									
REFERENCE	1 (bases 1 to 1081)									
AUTHORS	NIH-MGC	http://mgc.nci.nih.gov/								
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)									
JOURNAL	Unpublished (1999)									

COMMENT	Contact: Robert Strausberg, Ph.D. Email: cga@bse-mail.nih.gov Tissue Procurement: Gilbert Smith, Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM8560 row: b column: 04 High quality sequence stop: 641. Location/Qualifiers 1..1081 /organism="Mus musculus" /mol_type="mRNA" /strain="Czech II" /db_xref="taxon:10090" /clone="IMAGE:3500835" /tissue_type="spontaneous tumor, metastatic to mammary." Stem cell origin." /lab_host="DH10B" /clone_lib="NCI CGAP Lu29" /note="Organ: lung; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"
FEATURES	source
ORIGIN	Query Match 52.3%; Score 576; DB 2; Length 1081; Best Local Similarity 90.3%; Pred. No. 1.6e-113; Mismatches 65; Indels 4; Gaps 2; Matches 639; Conservative 0;
QY	388 GGTACAGTGCAC TGAAGTGTCTGGGAGCGGACGATCCGGGAAACGGAAGAGTCTTCA 447
Db	63 GGACCAAGTGCAC TGAAGTGTCTGGGAGTGCGAGCATCCGGGAAACGGAAGAGTCTTCA 122
QY	448 CAGAGCTCCGCCAGCTCGGAGAGAGAGTCCGCCCTGGATGAGATCATGAGGCTC 507
Db	123 CAGAGCTCCGCCAGCTCGGAGAGAGAGTCCGCCCTGGATGAGATCATGAGGCTC 182
QY	508 GAAGAGGAAAGAAAGGACCGGACGACGACGCTGTTACAGCCGGGGATCGTTGTG 567
Db	183 GAAGAGGAAAGAAAGGACCGGACGACGACGCTGTTACAGCCGGGGATCGTTGTG 242
QY	568 AAAATTTATAACGAAGAAGCTTTGGGAGAAATATATCAAGAAGAAAGGGTCTTAAAGGAA 627
Db	243 AAAATTTATAACGAAGAAGCTTTGGGAGAAATATATCAAGAAGAAAGGGTCTTAAAGGAA 302
QY	628 GTGATTGACAGGTACACAGCTGTGTGTTAAAGATGACTGACTCTGTGAGACAGGCTGAAACTG 687
Db	303 GTGATTGACAGGTACACAGCTGTGTGTTAAAGTGAAGT

```

663 TGACTTAACAAATAGGTCATCTATTACATCTTGTGGAGTAGTGGTAACCTTATTA- 721
1046 TAAAACTATGTTTTTTTAAAGTGTAAAAAAGAAAAAAGAAAAA 1093
722 -TAAACCTATGTTGTTAAACCGAAACCAACCAAAATATAGGAAAAA 768

RESULT 3
BG072126/c
LOCUS
DEFINITION
H3107A09-3 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
H3107A09 3', mRNA sequence.
ACCESSION
BG072126 GI:40072826
VERSION
BQ782294
KEYWORDS
SOURCE
MUS musculus (house mouse)
ORGANISM
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidae; Muridae; Murinae; Mus.
REFERENCE
Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X.,
Grahovac,M.J., Pantano,S., Sano,Y., Piao,Y., Nagaraja,R., Doi,H.,
Wood,W.H. III, Becker,K.G. and Ko,M.S.H.
Genome-wide expression profiling of mid-gestation placenta and
embryo using a 15,000 mouse developmental cDNA microarray
Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
10922068
On Jan 26, 2001 this sequence version replaced gi:12554695.
Other ESTs: H3107A09-5
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@gaun.grc.nia.nih.gov
This clone set has been freely distributed to the community. Please
visit http://igun.grc.nia.nih.gov/cDNA/15k.html for details.
Plate: H3107 row: A column: 09
Seq primer: -21M13 Forward
High quality sequence stop: 560
POLYA=Yes.
Location/Qualifiers
1..560
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="niaEST:H3107A09-3"
/db_xref="taxon:10090"
/clones="H3107A09"
/sex="Clones arrayed from a variety of cDNA libraries"
/dev_stage="Clones arrayed from a variety of cDNA
libraries"
/lab_host="DH10B"
/clone_lib="NIA Mouse 15K cDNA Clone Set"
/notes="vector: pSFR1; Site1: SalI; Site2: NotI; This
clone is among a rearranged set of 15,247 clones from 11
embryo cDNA libraries (including preimplantation stage
embryos from unfertilized egg to blastocyst, embryonic
part of E7.5 embryos, extraembryonic part of E7.5
embryos, and E12.5 female mesonephros/gonad) and one
newborn ovary cDNA library. Average insert size 1.5 kb.
All source libraries are cloned unidirectionally with
Oligo(dT)-Not primers. References include: (1)
Genome-wide expression profiling of mid-gestation
placenta and embryo using a 15,000 mouse developmental
cDNA microarray, 2000, Proc. Natl. Acad. Sci. U.S.A. 97:
9127-9132; (2) Large-scale cDNA analysis reveals phased
gene expression patterns during preimplantation mouse
development, 2000, Development, 127: 1737-1749; (3)
Genome-wide mapping of unselected transcripts from
extraembryonic tissue of 7.5-day mouse embryos reveals
enrichment in the t-complex and under-representation on

```

```

the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

ORIGIN
Query Match 50.7%; Score 559; DB 2; Length 560;
Best Local Similarity 100.0%; Pred. No. 7.1e-110; Indels 0; Gaps 0;
Matches 559; Conservative 0; Mismatches 0;

QY 512 AGGAAAGAAAAAGGACCGCACGACAGACGCTGTGTACAGCCGGGGATCGTTGTGAAAA 571
DB 560 AGGAAAGAAAAAGGACCGCACGACAGACGCTGTGTGTACAGCCGGGGATCGTTGTGAAAA 501
QY 572 TTATACGGAAGAGCTTGGGGAGAAATATACAGAGAGAAAGGGTTCGTTAGGAAGTGA 631
DB 500 TTATACGGAAGAGCTTGGGGAGAAATATACAGAGAGAAAGGGTTCGTTAGGAAGTGA 441
QY 632 TTGACAGGTACACAGCTGTGTAAGATGACTCTCTGAGACAGCGCTGAAACTCGACC 691
DB 440 TTGACAGGTACACAGCTGTGTAAGATGACTCTCTGAGACAGCGCTGAAACTCGACC 381
QY 692 AGACTCATTTAGACAGCAGTCATTCCGGCCCCCGGGAAAAAGGGTTCCTAGTTTAAATGGAG 751
DB 380 AGACTCATTTAGACAGCAGTCATTCCGGCCCCCGGGAAAAAGGGTTCCTAGTTTAAATGGAG 321
QY 752 GTTACAGAGAAATGAAGGCACCTCTCGAATCCATCAATGAGAAGGCTTTTTCAGCCACGA 811
DB 320 GTTACAGAGAAATGAAGGCACCTCTCGAATCCATCAATGAGAAGGCTTTTTCAGCCACGA 261
QY 812 TAGTCATTGAAACTCGACCTTTGAAAGGACGAGAGTTCGAAGGTATTCAATATGAAGACA 871
DB 260 TAGTCATTGAAACTCGACCTTTGAAAGGACGAGAGTTCGAAGGTATTCAATATGAAGACA 201
QY 872 TATCTAAACTTGCTTGAGTTTGAAAAATTTGATAACACACATTTGAAACTGTGAAGCATCA 931
DB 200 TATCTAAACTTGCTTGAGTTTGAAAAATTTGATAACACACATTTGAAACTGTGAAGCATCA 141
QY 932 AATTGCTGTAGCAAGGCACCTGTCTACTCTGTGTAGGGGATTTGTTTGTATTATTA 991
DB 140 AATTGCTGTAGCAAGGCACCTGTCTACTCTGTGTAGGGGATTTGTTTGTATTATTA 81
QY 992 AAAAAAANAATCATCTATTTAAATACTAGTGAATAGTTGGGTAAATTTTATATAAAT 1051
DB 80 AAAAAAANAATCATCTATTTAAATACTAGTGAATAGTTGGGTAAATTTTATATAAAT 21
QY 1052 CTATGTTTTTTTTTAAGTGT 1070
DB 20 CTATGTTTTTTTTTAAGTGT 2

RESULT 4
BQ782294/c
LOCUS
DEFINITION
UI-R-FF0-cpj-b-08-0-UI.s1 NCI CGAP_FF0 Rattus norvegicus cDNA clone
UI-R-FF0-cpj-b-08-0-UI 3', mRNA sequence.
ACCESSION
BQ782294
VERSION
BQ782294.1 GI:21990766
KEYWORDS
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidae; Muridae; Murinae; Rattus.
REFERENCE
1 (bases 1 to 739)
AUTHORS
Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL
Genome Res. 6 (9), 791-806 (1996)
PUBMED
8889548
COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565

```

Email: bento-soares@uiowa.edu
 Tissue Procurement: Jeff Stevens
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: DISTRIBUTION: Researchers may obtain clones
 from Research Genetics (www.resgen.com).
 The following repetitive elements were found in this cDNA
 sequence: 1-47, >AT-rich#Low_complexity (matched complement)
 Seq primer: M13 FORWARD
 POLYA=Yes.

FEATURES
 source
 Location/Qualifiers
 1..739
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /db_xref="taxon:10116"
 /clone="UI-R-PF0-cpj-b-08-0-UI"
 /tissue_type="Mixed tissues"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_lib="NCI CGAP PF0"
 /note="Vector: pT7T3-Pac (Pharmacia) with a modified
 polylinker; Site 1: EcoR I; Site 2: Not I; UI-R-PF0 is a
 subtracted cDNA library containing the following
 tissue(s): Normal cartilage and SR-JWS Tumor Line . The
 subtraction was made according to Bonaldo, Lennon and
 Soares, Genome Research, 6:791-806, 1996. The
 oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tags for these libraries are: CTAATGGAGC,
 CATTCTTGTA.
 TAG_TISSUE=rat SRC-JWST tumor line
 TAG_LIB=UI-R-PF0
 TAG_SEQ=CATTCTTGTA"

ORIGIN

Query Match 50.3%; Score 554.2; DB 5; Length 739;
 Best Local Similarity 90.2%; Pred. No. 7.9e-109;
 Matches 632; Conservative 0; Mismatches 53; Indels 16; Gaps 3;

QY 388 GTTACAGTGCCTGAAGCTGCTGGGAGCGAGCATCCGGAAACGGAAGAGCTTCA 447
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 692 GGACCGAGCGCTGAAGATGCTGGGAGCAGCATCCGTGAGCGGAAGAGCTTCA 633
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 448 CAGAGCTCCGCCAGCTGC---GAAGAAGAAGAGTCCGCCCTGGATGAGATCATGGAG 504
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 632 CAGAGCTCAGCTCAGCTCGGAAGAAGAAGAGTCTGCCCTGGATCAGATCATGGAG 573
 QY 505 CTCGAAGAGAAAGAAAGAGCCGACGACGACGCTGTTACGCGGGGATCGTT 564
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 572 ATTGAGGAGAAAGAAAGAGCCGACGACGACGACTCTCTGGTTACAGCTCGGATCAT 513
 QY 565 GTGAAATTTATACGAAGAGCTTGGGAGAAATATCACAGAAGAAAGGGTCGTTAG 624
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 512 GTGAAATTTATACGAAGAGCTTGGGAGAAATATCACAGAAGAAAGGGTCGTTAG 453
 QY 625 GAAGTGATTGACAGGTACACAGCTGTGGTAAAGATGACTCTCTGGACACAGGCTGAA 684
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 452 GAAGTGATTGACAGATACACAGCTGTGGTAAAGATGACTCTCTGGACACAGGCTGAA 393
 QY 685 CTGGACCAAGCTATTAGAGACAGTCATCTCGGCCCGCGGAAAGAGGTTCTAGTTTA 744
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 392 CTGGACCAAGCTATTAGAGACAGTCATCTCGGCCCGCGGAAAGAGGTTCTAGTTTA 333
 QY 745 AATGAGGCTACAGAGGAATGAAGCACTCTCGATCCATCAATGAGAGGCTTTTCA 804
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 332 AACGAGGCTACAGAGGAATGAAGCACTCTCGATCCATCAATGAAAGACTTTTCA 273
 QY 805 GCCAGATAGTCATTGAAACTGGACCTTTTGAAGGACGAGAGTTGAAGGTTTCAATAT 864
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 272 GCCAGATAGTCATTGAAACGGACCTTTTGAAGGCGCAGACTTGAAGGTTTCAATAT 213

QY 865 GAAGACATATCTAAACTTGCCTTGAGTTTGAAATTTGATAACACACACATTTGAACCTGCA 924
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 212 GAAGACATATCTAAACTTGCCTTGAGTTTGAAATTTGATAACACACATTTGAATTTGTA 153
 QY 925 AGCATCAAAATTTGGTGTAGCCAAAGCAGCTGTGTAACTCTACTGTCTAGGGGATTTGTTT 984
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 152 AGCATCAAAATTTGAAGTTAGCCAGCAGCTGTGTAACTCTACTGTCTAGGGGATTTGTTT 93
 QY 985 TGTATTAAAAAATAAATCATCTATTAAATACTAGTGAATAG-----TTGGGTAATTT 1040
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 92 -----ATAAGAAAGTCATCTATTAAATACCACTGGATAGCTGTTTGGTAAATTT 42
 QY 1041 TATAATAAATCTATGTTTTTTTAAAGTGTAAAAAATAAATA 1081
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 41 TATAATAAATCTAATTTTTTTTTTAAAAAATAAATAAATAAATA 1

RESULT 5
 BQ780218/c
 LOCUS
 DEFINITION
 BQ780218 754 bp mRNA linear EST 26-JUL-2002
 UI-R-PF0-cpb-m-06-0-UI-s1 NCI CGAP PF0 Rattus norvegicus cDNA clone
 UI-R-PF0-cpb-m-06-0-UI 3', mRNA sequence.
 BQ780218
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Rattus norvegicus (Norway rat)
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridea; Muridae; Murinae; Rattus.
 1 (bases 1 to 754)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 8889548
 Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 Tissue Procurement: Jeff Stevens
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: DISTRIBUTION: Researchers may obtain clones
 from Research Genetics (www.resgen.com).
 The following repetitive elements were found in this cDNA
 sequence: 1-54, >AT-rich#Low_complexity (matched complement)
 Seq primer: M13 FORWARD
 POLYA=Yes.

FEATURES
 source
 Location/Qualifiers
 1..754
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /db_xref="taxon:10116"
 /clone="UI-R-PF0-cpb-m-06-0-UI"
 /tissue_type="Mixed tissues"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_lib="NCI CGAP PF0"
 /note="Vector: pT7T3-Pac (Pharmacia) with a modified
 polylinker; Site 1: EcoR I; Site 2: Not I; UI-R-PF0 is a
 subtracted cDNA library containing the following
 tissue(s): Normal cartilage and SR-JWS Tumor Line . The
 subtraction was made according to Bonaldo, Lennon and
 Soares, Genome Research, 6:791-806, 1996. The
 oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tags for these libraries are: CTAATGGAGC,

RESULT 7	BM228037/c	512 bp	mRNA	linear	EST 07-JUN-2003
LOCUS	K0254C06-3	NIA Mouse Unfertilized Egg cDNA Library (Long)	Mus		
DEFINITION	musculus cDNA clone NIA:K0254C06 IMAGE:30049853 3', mRNA sequence.				
ACCESSION	BM228037				
VERSION	BM228037.2	GI:31485640			
KEYWORDS	EST.				
SOURCE	Mus musculus	(house mouse)			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	Piao,Y., Ko,N.T., Lim,M.K. and Ko,M.S.H.				
AUTHORS	1 (bases 1 to 512)				
TITLE	Construction of long-transcript enriched cDNA libraries from submicrogram amounts of total RNAs by a universal PCR amplification method				
JOURNAL	Genome Res.	11 (9)	1553-1558	(2001)	
PUBMED	11544199				
COMMENT	On Dec 14, 2001 this sequence version replaced gi:17790538. Other ESTs: K0254C06-5N Contact: Dawood B. Dudekula Laboratory of Genetics National Institute on Aging/National Institutes of Health 333 Casseell Drive, Suite 4000, Baltimore, MD 21224-6820, USA Email: cdna@lgun.grc.nia.nih.gov Plate: K0254 row: C column: 06 Seq primer: -21M13 Forward High quality sequence stop: 512 POLYA=Yes.				
FEATURES	Location/Qualifiers				
source	1..512				
	/organism="Mus musculus"				
	/mol_type="mRNA"				
	/strain="C57BL/6J"				
	/db_xref="niaEST:K0254C06-3"				
	/db_xref="taxon:10090"				
	/clone="NIA:K0254C06 IMAGE:30049853"				
	/tissue_type="Unfertilized Egg"				
	/lab_host="DH10B"				
	/clone_lib="NIA Mouse Unfertilized Egg cDNA Library (Long)"				
	/notes="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://lgun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were extracted from a pool of 1488 unfertilized eggs. Double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen].				
	5'-pGACTAGTCTAGATCGGAGCGGCCCTTTTTTTTTTTT-3'], treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to Lone-linker LI-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.5 kb. The library was constructed by Yulan Piao (NIA)."				
ORIGIN	Query Match 46.5%; Score 512; DB 3; Length 512; Best Local Similarity 100.0%; Pred. No. 9.7e-100; Indels 0; Matches 512; Conservative 0; Mismatches 0; Gaps 0;				
	479 AGTCGGCCCTGGATGAGATCATGAGCTCGAAGAGAAAGAAAGACCGCAGCAG 538				

```

/db xref="taxon:10090"
/clone="NIA:K0278H05 IMAGE:30052216"
/tissue_type="Unfertilized Egg"
/lab_host="DH10B"
/clone_lib="NIA Mouse Unfertilized Egg cDNA Library
(Long)"
/notes="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
extracted from a pool of 1488 unfertilized eggs.
Double-stranded cDNAs were synthesized with an Oligo(dT)
primer [Invitrogen]:
5'-pGACTAGTCTTAGATCCGAGCGGCCCTCTTTT-3',
treated with T4 DNA polymerase, and purified by
ethanol-precipitation. The cDNAs were ligated to
Lone-linker L1-Sal4, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer Sal4-S. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pSPORT1 plasmid vector.
The DH10B E. coli host was transformed with the ligation
mixture by the standard chemical method. The average
insert size is about 2.5 kb. The library was constructed
by Yulan Piao (NIA)."
```

ORIGIN

```

Query Match      46.5%; Score 512; DB 3; Length 512;
Best Local Similarity 100.0%; Pred. No. 9.7e-100;
Matches 512; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 479 AGTCGGCCCTGGATGAGATCATGAGCTCGAAGAGGAGAAAGAAAGCAGCGACAG 538
Db 512 AGTCGGCCCTGGATGAGATCATGAGCTCGAAGAGGAGAAAGAAAGCAGCGACAG 453
Qy 539 ACGCTGTGTACAGCCGGGGATCGTTGTGAAATATTAACGAAGAAGCTTGGGAGAAAT 598
Db 452 ACGCTGTGTACAGCCGGGGATCGTTGTGAAATATTAACGAAGAAGCTTGGGAGAAAT 393
Qy 599 ATCAAGAAGAAAGGGTCTGTTAAGGAAGTATTGACAGTACACAGCTGTGTTAAGA 658
Db 392 ATCAAGAAGAAAGGGTCTGTTAAGGAAGTATTGACAGTACACAGCTGTGTTAAGA 333
Qy 659 TGACTGACTCTGGAGACAGGCTGAAACTGGACACAGACTCATTTAGAGACAGTCAATCCGG 718
Db 332 TGACTGACTCTGGAGACAGGCTGAACTGGACACAGACTCATTTAGAGACAGTCAATCCGG 273
Qy 719 CCCCAGGAGAAAGGGTCTAGTCTTTAAATGAGGCTACAGAGGAATGAAGGCACTCTCG 778
Db 272 CCCCAGGAGAAAGGGTCTAGTCTTTAAATGAGGCTACAGAGGAATGAAGGCACTCTCG 213
Qy 779 AATCATCAATGAGAGGCTTTTTCAGCCAGTATGATCATTTGAACTGGACCTTTGAAG 838
Db 212 AATCATCAATGAGAGGCTTTTTCAGCCAGTATGATCATTTGAACTGGACCTTTGAAG 153
Qy 839 GAGCAGAGTTTGAAGGTATTCAATATGAAGACATATCTAAATCTTGTAGTTTGAAGAT 898
Db 152 GAGCAGAGTTTGAAGGTATTCAATATGAAGACATATCTAAATCTTGTAGTTTGAAGAT 93
Qy 899 TTGATAACAACAATTTGAAACTGTGAAGCATCAAAATTTGGTGTAGCCAGGCACTGTGTA 958
Db 92 TTGATAACAACAATTTGAAACTGTGAAGCATCAAAATTTGGTGTAGCCAGGCACTGTGTA 33
Qy 959 ACTCTACTGTGTAGGGGATTTGTTTGTATT 990
Db 32 ACTCTACTGTGTAGGGGATTTGTTTGTATT 1
```

RESULT 9

```

AY412520
LOCUS      1154 bp      DNA      linear      GSS 16-DEC-2003
DEFINITION Mus musculus KIN gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION  AY412520
VERSION     AY412520.1 GI:39768485
KEYWORDS   GSS.
SOURCE      Mus musculus (house mouse)
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 1154)
AUTHORS     Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Perrieri,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
            Adams,M.D. and Cargill,M.
            Direct Submission
            Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
            This sequence was made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES             Location/Qualifiers
     source           1..1154
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     <1..>1154
                     /genes="KIN"
                     /locus_tag="HCM4584"
     ORIGIN
Query Match      45.0%; Score 496.2; DB 10; Length 1154;
Best Local Similarity 99.4%; Pred. No. 2.7e-96;
Matches 498; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 388 GGTACCAAGTGCACCTGAAGCTGCTGGGAGCGGAGCATCCGGGAAACGGAAGAGTCTTCA 447
Db 654 GGACCAAGTGCACCTGAAGCTGCTGGGAGCGGAGCATCCGGGAAACGGAAGAGTCTTCA 713
Qy 448 CAGAGCTCCGCCAGCTCGGAAGAGAAAGAAAGTCCGCCCTCGATGAGATCATGGAGCTC 507
Db 714 CAGAGCTCCGCCAGCTCGGAAGAGAAAGAAAGTCCGCCCTCGATGAGATCATGGAGCTC 773
Qy 508 GAAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 567
Db 774 GAAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 833
Qy 568 AAAATTATTAACGAAGAGCTTGGGAGAGAAATATCACAGAGAAAGAAAGAAAGAAAGAAAG 627
Db 834 AAAATTATTAACGAAGAGCTTGGGAGAGAAATATCACAGAGAAAGAAAGAAAGAAAGAAAG 893
Qy 628 GTGATTACAGAGGTACACAGCTGTGTTAAAGATGACTGACTCTCGAGACAGCGTGAAGAACTG 687
Db 894 GTGATTACAGAGGTACACAGCTGTGTTAAAGATGACTGACTCTCGAGACAGCGTGAAGAACTG 953
Qy 688 GACCAGACTCATTTAGAGACAGTCAATTCGGCCCCCGGGGAAAGGGTCTTAGTTTAAAT 747
Db 954 GACCAGACTCATTTAGAGACAGTCAATTCGGCCCCCGGGGAAAGGGTCTTAGTTTAAAT 1013
Qy 748 GGAGGCTACAGAGAAATGAAGGCACTCTCGAATCCATCAATGAGAGGCTTTTTCAGCC 807
Db 1014 GGAGGCTACAGAGAAATGAAGGCACTCTCGAATCCATCAATGAGAGGCTTTTTCAGCC 1073
```

QY 808 ACATAGTTCATTGAACCTGGACCTTTGAAAGGACGCGAGATTGAAGGTATTCAATATGAA 867
|||||
Db 1074 ACATAGTTCATTGAACCTGGACCTTTGAAAGGACGCGAGATTGAAGGTATTCAATATGAA 1133
|||||
QY 868 GACATATCTAAACTTGCCTTGA 888
|||||
Db 1134 GACATATCTAAACTTGCCTTGA 1154
|||||

RESULT 10
LOCUS CR595908
DEFINITION full-length cDNA clone CS0D015Y119 of T cells (Jurkat cell line)
ACCESSION CR595908
VERSION CR595908.1 GI:50476715
KEYWORDS HTC; CNSLT cdNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1396)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
TITLES Li, W. B., Gruber, C., Jessee, J., and Polayes, D.
JOURNAL Full-length cDNA libraries and normalization
REMARK Unpublished
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paraday Avenue
2 (bases 1 to 1396)
REFERENCE Genoscope.
AUTHORS Direct Submission
TITLES Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.

FEATURES
source
1..1396
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D015Y119"
/tissue_type="T cells (Jurkat cell line) Cot
10-normalized"
/plasmid="pCMVSPORT_6"

ORIGIN
Query Match 43.0%; Score 474; DB 4; Length 1396;
Best Local Similarity 65.4%; Pred. No. 1.7e-91;
Matches 896; Conservative 0; Mismatches 180; Indels 294; Gaps 2;
QY 2 TGGGCAAGTCGGAATTTCTGAGCCCCCAAGGCCATCGCCCAATGAAATTAAGTCCAAAGGCG 61
|||||
Db 1 TGGGCAAGTCGGAATTTCTTACTCCCAAGGCTATCGCCCAAGGATCAAGTCCAAAGGCGC 60
|||||
QY 62 TCCAGAACTCGCTGGTACTCCAGATGTGCCAAAGCAATGCCGCGACGAGATGGCT 121
|||||
Db 61 TCCAGAACTCGCTGGTATGCGAGATGTGCCAAAGCAAGTGTGCCGGGACGAGATGGCT 120
|||||
QY 122 TTAAGTGTCACTGTATGTCTGAATCTCATCAAGACAACATGTTGCTGGCTTCAGAAAACC 181
|||||
Db 121 TTAAGTGTCACTGTATGTCCGAATCTCATCAGAGACAACATGTTGCTGGCTTCAGAAATC 180
|||||
QY 182 CTCAGCAGTTTATGGATTATTTTCAGAGGAATTCGGAATGACTTCTTGGAACTTCTGA 241
|||||
Db 181 CTCAGCAGTTTATGGATTATTTTCAGAGGAATTCGGAATGACTTCTTGAACCTTCTCA 240
|||||
QY 242 GCGAGCTTTGGCACTAAAGGCTCCACAAACATCTGTACAAATGAATACATCAGCC 301
|||||

Db 241 GGAGAGCGTTTGGCACTAAAAGGCTCCACAACACATTTGTCTACAAACGAATACATCAGCC 300
|||||
QY 302 ACCGAGACCATCCATCAATGAACCGCTACCCAGTGGGAGACACTGACCCGACTTTTACCAAGT 361
|||||
Db 301 ACCGAGACCATCCATCAATGAATGCCACTCAGTGGGAACTCTGACTGATTTTACTAAGT 360
|||||
QY 362 GGCTGGGCAGAGAGGGCTTGTGTAAA----- 387
|||||
Db 361 GGCTGGGCAGAGAGGGCTTGTGCAAAAGTGGACGAGACACCAAAAGGCTGGTATATTTCAGT 420
|||||
QY 388 ----- 387
|||||
Db 421 ACATAGACAGGACCCAGAAAACATATCCCGCGCAACTGGAACCTGGAGAAAAAGAAAAGC 480
|||||
QY 388 ----- 387
|||||
Db 481 AGGACCTTTGATGATGAAGAAAAAAACTGCCCCAAATTTATTGAAGAGCAAGTGAGAGAGGCC 540
|||||
QY 388 ----- 387
|||||
Db 541 TGAAGGGAAGAAACAGAGAGTCCCTACTTTTACGGAATTAAGCAGAGAAAAATGATGAAG 600
|||||
QY 388 ----- 387
|||||
Db 601 AGAAAGTCACGTTTAATTTGAGTAAAGGAGCATGTAGCTCATCCGGAGCAACATCTTCCA 660
|||||
QY 388 -----GGTACCAGTGCATGAAGCTGCTGGGAGGCGCAGCATCCGGGAAAC 433
|||||
Db 661 AGTCAAGTACTCTGGGACCGAGTGCCTGAAAGCATAGGAAGTTTCAGCATCAGTGAAAC 720
|||||
QY 434 GGAAGAGTCTTCACAGAGCTCCGCCAGCTGC-----GAGAGAGAGAGTCCGCC 487
|||||
Db 721 GAAAGAAATCTTCCAGAGCTCAACTCTAGCTTAAAGAAAAGAAAAGAAATCTGCAC 780
|||||
QY 488 TGGATGAGATCATGAGGCTCGAAGAGGAAAAAGAACCGCACGACAGACGCTGCT 547
|||||
Db 781 TGGATGAATCATGAGGATTTGAGAGGAAAAGAAAGAACTGCCCGAACAGACTACTGCG 840
|||||
QY 548 TACAGCCGGGATCGTTGTGAAAAATTATAACGAAGAACTTGGGGAGAAAATATCACAGA 607
|||||
Db 841 TACAGCTCTGAAATTTATGTGAAAAATTATAACGAAGAACTTGGGAGAGAAAATATACATA 900
|||||
QY 608 AGAAAGGCGTCGTTAAGGAAGTGATTGACAGGTACACAGCTGTGTTAAGATGACTGACT 667
|||||
Db 901 AAAAGGCTATTGTTAAGGAAGTAAATTGACAAATATACAGCTGTTGTGGAAGATGATTGAT 960
|||||
QY 668 CTGGAGACAGGCTGAAACTGGACACAGACTCATTTAGAGACAGTCAATCCGGCCCCGGGGA 727
|||||
Db 961 CTGGAGACAGGCTGAAACTTGACACAGACTCATTTAGAGACAGTAAATTCAGCACCAGAA 1020
|||||
QY 728 AAAGGTTCTAGTTTTAAATCGAGGCTACAGAGGAAATGAAGGCACTCTCGAATCCATCA 787
|||||
Db 1021 AAAGAAATCTAGTTTTAAATGGAGGCTACAGAGGAAATGAAGGTACCTAGAAATCCATCA 1080
|||||
QY 788 ATGAGAAGGCTTTTTCAGCCACGATAGTCAATGAACTGGACCTTTGAAAGACCCAGAG 847
|||||
Db 1081 ATGAGAAGGCTTTTTCAGCTACTATCGTCAATGAACTGGCCCTTTAAAGACCCAGAG 1140
|||||
QY 848 TTGAAGGTATTCAATATGAAGACATATCTAAACTTGTGAGTTTGAATAATAACA 907
|||||
Db 1141 TTGAAGGAATTCATATGAAGACATTTCTAAACTTGCCTGAGTTTGAATAATAACA 1200
|||||
QY 908 ACACATTTGAACTGTGAAGCATCAAAATTTGGTGTGTAGCCAAAGGCACTGTGTAACTCTACTG 967
|||||
Db 1201 ATACATTTAAATCTTTAAAGCATCAAAATTTGGTGTTCGCCAAGGCATTATGAGACTCTACTG 1260
|||||
QY 968 TGTTAGGGGAATTTGTTTGTATTAAAAAAAATAAATCATCTATTTTAAATACTAGTGAAT 1027
|||||
Db 1261 TGTTAGGGGTATTTCTTTTGTATAAAAACAACAGGTTTTTGAAAATATTTACTGTATAGTT 1320
|||||
QY 1028 AGTTGGGTAAATTTTATAATAAAATCTATGTTTTTTTTTTTAAAGTGTAAAAAAA 1077
|||||
Db 1321 GTTCAGCTTAACCTTTGAGAGAAATTTAATTATATGCTCATGAGGTATCAAA 1370
|||||

provided by Dr. Bertrand Jordan. Library went through three rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN	Query Match Best Local Similarity Matches 466; Conservative 0; Mismatches 0; Indels 2; Gaps 1;
QY	529 GCACGACAGACGCGCTGGTTACAGCGCGGGATCGTTGTGAAATATATACGAGGAAGCTT 588
Db	1 GCACGACAGACGCGCTGGTTACAGCGCGGGATCGTTGTGAAATATATACGAGGAAGCTT 60
QY	589 GGGGAGAAATATACAGAGAAAGAGGGTCGTTAAGGAAGTGAATGACAGGTACACAGCT 648
Db	61 GGGGAGAAATATACAGAAG--AAAGGGTCGTTAAGGAAGTGAATGACAGGTACACAGCT 118
QY	649 GTGGTAAAGATCACTGACTCTCGAGACAGGCTGAACTGGACACAGCTCATTTAGAGACA 708
Db	119 GTGGTAAAGATCACTGACTCTCGAGACAGGCTGAACTGGACACAGCTCATTTAGAGACA 178
QY	709 GTCAATTCGGCCCCCGGGAAAGGGTCTAGTGTGTTAAATGGAGGCTACAGAGCAATGAA 768
Db	179 GTCAATTCGGCCCCCGGGAAAGGGTCTAGTGTGTTAAATGGAGGCTACAGAGCAATGNA 238
QY	769 GGCACCTCTCGAATCAATCAGAAAGGCTTTTCAGCCACGATGATGTAATGAACTGGA 828
Db	239 GGCACCTCTCGAATCAATCAGAAAGGCTTTTCAGCCACGATGATGTAATGAACTGGA 298
QY	829 CTTTGAAGGACGAGAGTCAAGGTATTCATATGAGACATATCTAACTCTCTTCA 888
Db	299 CTTTGAAGGACGAGAGTCAAGGTATTCATATGAGACATATCTAACTCTCTTCA 358
QY	889 GTTGAATTTGATAACCAACATTTGAACTGTGAAAGCATCAATTTGGTTAGCCAAAG 948
Db	359 GTTGAATTTGATAACCAACATTTGAACTGTGAAAGCATCAATTTGGTTAGCCAAAG 418
QY	949 GCACGTGTGTAACCTACTGTGTAGGGGATTTGTTTGTATTAATAA 996
Db	419 GCACGTGTGTAACCTACTGTGTAGGGGATTTGTTTGTATTAATAA 466
RESULT 13	
AY609991	1543 bp mRNA linear HTC 31-JAN-2005
DEFINITION	Sus scrofa clone Clu_5302.scr.msk.pl.Contig4, mRNA sequence.
ACCESSION	AY609991
VERSION	AY609991.1 GI:52351561
KEYWORDS	HTC.
SOURCE	Sus scrofa (pig)
ORGANISM	Sus scrofa
REFERENCE	Jorgensen,F.G., Hobolth,A., Hornshoj,H., Bendixen,C., Fredholm,M. and Schierup,M.H.
AUTHORS	Comparative analysis of protein coding sequences from human, mouse and the domesticated pig
TITLE	(er) BMC Biol. 3 (1), 2 (2005)
JOURNAL	15679890
PUBMED	2 (bases 1 to 1543)
REFERENCE	Hornshoj,H., Bendixen,C. and Panitz,F.
AUTHORS	Submitted (28-APR-2004) Animal Breeding and Genetics, Danish
TITLE	Institute of Agricultural Sciences, Research Centre Foulum, Postbox
JOURNAL	50, Tjele DK-8830, Denmark
FEATURES	Location/Qualifiers
source	1..1543
	/organism="Sus scrofa"
	/mol_type="mRNA"
	/db_xref="taxon:9823"
	/clone="Clu_5302.scr.msk.pl.Contig4"

misc_feature	1..1543	/notes="similar to NM_012311.2 Homo sapiens KIN, antigenic determinant of recA protein homolog (mouse) (KIN)"	
ORIGIN			
Query Match	40.7%;	Score 449;	DB 4; Length 1543;
Best Local Similarity	65.6%;	Pred. No. 4.2e-86;	
Matches 871; Conservative	0;	Mismatches 160;	Indels 296; Gaps 3;
Qy	1	ATGGCAAGTCGGATTTTCTGAGCCCCAAGGCCATCGCCAATAGAAATTAAGTCCAAAGGG	60
Db	36	ATGGCAAGTCGGATTTTCTTACCCGAAGGCCATCGCCAACAGGATCAAAATCCAAAGGG	95
Qy	61	CTCCAGAGCTTCGCTGTGACTGCCAGATGTGCCAAAGCAATGCCGCGACGAGAAATGGC	120
Db	96	CTTCAGAAAGCTCGCGCTGGTATTGCCAGATGTGCCAAAGCAGTGCCTGGGACGAGAAATGGC	155
Qy	121	TTTAAAGTGTCACTGTATGCTGAATCTCATCAAAAGACAACTGTTGCTGCTTCAGAAAAAC	180
Db	156	TTTAAAGTGTCACTGTATGCTCGAATCGCATCAGACAACTGTTGCTGCTTCGAAAT	215
Qy	181	CCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTCGAAATGACTTCTCGAACTTCTG	240
Db	216	CCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTCGAAATGACTTCTTAGAATCTCTC	275
Qy	241	AGGCGACGCTTTGGCACTAAAGGGTCCAAACAATTTGCTTACAATATACATCAGC	300
Db	276	AGGAGACGCTTTGGCACCAAGAGGGTTCAACAACAATCGTCTACAACGAGTACATCAGC	335
Qy	301	CACCGAGACATCCACATGAACGCTTACCAGTGGGAGACATGTACCGACTTTTACCAG	360
Db	336	CACCGAGACATCCACATGAACGCTTACCAGTGGGAGACGCTGACGGATTTTCCACCA	395
Qy	361	TGCGTGGCGAGAGGGCTTGTGTAAA-----	387
Db	396	TGCGTGGCGAGAGAGGCTTGTGCAAGTGGATGAGACACAAAGGCTGGTATATTTCAG	455
Qy	388	-----	387
Db	456	TACATAGATAGGACCCCGGAAACAATCCGCGCGCACTGGAATAGAGAAAAAGAGAAG	515
Qy	388	-----	387
Db	516	CAGGACTTGGACGATGAAGAAAAAACTGCCAAATTTATTCAAGAAACAAGTGAGACGAG	575
Qy	388	-----	387
Db	576	CTGGAAGGGAAGGAGCAGGAGGCCCTTATTTTACGGAGTTAAGCAGAGAAAAATGAAGAA	635
Qy	388	-----	387
Db	636	GAAAAAGTGACATTTAAATTTGAAACAAAGGAGCATGTAGTTTCAGCAGCAGCATCTTCAA	695
Qy	388	-----GGTACCACTGCACCTGAAGCTGCTGGGAGCGCAGCATCCGGGAAACGG	435
Db	696	TCAAGTTCTTTGGACCAAGTGCCCTGAAGACGATGGGTAGCAGCGCTCGGTGAACGA	755
Qy	436	AAAGAGCTTTCACAGAGCTCCGCCAGCGCTGC-----GAAGAAGAAGAAGTCGCGCCTG	489
Db	756	AAGGAATCTTCCAGAGCTCAGCTCAGTCAAAAGAAAAAGAAAAAGAGTCTCGCGCTC	815
Qy	490	GATGAGATCATGAGCTCGAAGAGGAAGAAAGAAAGGACCGCACGGACAGACCGCTGGTTA	549
Db	816	GACGAAATCATGGAGATTGAAGAGGAGAAAGAGAACTCCCGGACGGATCACTGGCTC	875
Qy	550	CAGCCGGGATCGTGTGAAAAATTTATAACCAAGAGCTTCGGGAGAAATATCACAGAAG	609
Db	876	CAGCTGAAATTTATAGTGAATTTATAACCAAAACTTGGAGAGAAATATCATAGAA	935
Qy	610	AAAGGGGTTCGTTAAGGAAGTGAATGACAGGTAACACAGCTGGGTAAAGATGACTGCTCT	669
Db	936	AAGGCGATTTGTTAAGGAAGTAAATTGACAAATACACAGCTGTTGTTAAGATGATTGACTCT	995


```
QY 670 GGAGACAGCGTGAACCTGACACAGACTCATTTAGACAGACTCATTCGGCCCGGGGAAA 729
Db 996 GGAGACAGCGTGAACCTGACACAGACTCATTTAGACAGACTCATTCGGCCCGGGGAAA 1055
QY 730 AGGGTCTAGTTTAAATGAGGCTACAGAGGAATGAAGGCACTCTCGAATCCATCAAT 789
Db 1056 AGAATCTCGTCTGTAATGAGGCTACAGAGGAATGAAGGCACTCTCGAATCCATCAAC 1115
QY 790 GAGAAGCGTTTTCAGCCAGTAGTCAATGAACTGGACCTTTGAAGAGCGCAGAGTT 849
Db 1116 GAGAAGCGTTTTCAGCTACTAGTCAATGAACTGGACCTTTGAAGAGCGCAGAGTT 1175
QY 850 GAAGGTATTCAATATGAAGCATATCTAACTTGTCTGATTTGAAATTTGATAACAAC 909
Db 1176 GAAGGAATCAATATGAGACATTTCTAACTTGTCTGATTTGAAATTTGATAACAAC 1235
QY 910 ACATTGAACTGTGAAGCACTCAATATGTTGTTAGCCAGGCACTGTGTAACTTCTACTGTG 969
Db 1236 ACATTGAACTGTGAAGCACTCAATATGTTGTTAGCCAGGCACTGTGTAACTTCTACTGTG 1295
QY 970 TTAGGGGATTTGTTTGTATTAAAAAATAAATCATCTATTAAATACTAGTGAATAG 1029
Db 1296 TTAGGGGATTTGTTTGTATTAAAAAATAAATCATCTATTAAATACTAGTGAATAG 1350
QY 1030 TTGGGTA 1036
Db 1351 CTGTTTA 1357
```

```
RESULT 14
AW045702/c 457 bp mRNA linear EST 18-SEP-1999
LOCUS UI-M-BH1-akq-d-04-0-UI.s1 NIH_BMAP_M_S2 Mus.musculus cDNA clone
DEFINITION UI-M-BH1-akq-d-04-0-UI 3', mRNA sequence.
ACCESSION AW045702
VERSION AW045702.1 GI:5906231
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
```

```
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
TITLE Sciurognathi; Muroidae; Muridae; Murinae; Mus.
1 (bases 1 to 457)
Bonaldo,W.F., Lemmon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
8889548
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: m5t@mail.nih.gov
```

```
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized corpus striatum library cDNA Library Preparation: M.B.
Soares Lab Clone distribution: NIH BMAP cDNA clones will be made
available by the means that is soon to be determined. When NIH
determines the means for distribution of the BMAP cDNA clones, this
record will be updated accordingly when that means is determined.
```

```
The following repetitive elements were found in this cDNA sequence:
1-21, >AT-rich#Low_complexity
Seq primer: M13_Foward
POLYA=Yes.
Location/Qualifiers
1..457
/organism="Mus musculus"
/mol_type="mRNA"
```

FEATURES
source

```
/strain="C57BL/6J"  
/db_xref="taxon:10090"  
/clone="UI-M-BH1-akq-d-04-0-UI"  
/dev_stage="27-32 days"  
/lab_host="DH10B (Life Technologies)"  
/clone_lib="NIH_BMAP_M_S2"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; The  
NIH_BMAP_M_S2 library is a subtracted library derived from  
NIH_BMAP_M_S1, which in turn is a subtracted library  
derived from a mixture of normalized libraries from ten  
regions of the mouse brain (cerebellum, brain stems,  
olfactory bulbs, hypothalamus, cortex, amygdala, basal  
ganglia, pineal gland, striatum, hippocampus). The driver  
used for subtraction consisted of a pool of 5,000 clones  
from the NIH_BMAP_M_S1 library and a pool of 2,000 clones  
obtained from non-normalized and normalized mouse brain  
spinal cord libraries.  
TAG_TISSUE=corpus-striatum  
TAG_LIB=NIH_BMAP_M_S2  
TAG_SEQ=ACGGC"
```

ORIGIN

Query Match 40.7%; Score 448.2; DB 1; Length 457;
Best Local Similarity 99.3%; Pred. No. 5.Se-86;
Matches 450; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 551 AGCCGGGATCGTTGTGAATTTATTAACGAAGAGCTTCGGGAGAAATATCACAGAAGA 610
Db 457 AGCCGGGATCGTTGTGAATTTATTAACGAAGAGCTTCGGGAGAAATATCACAGAAGA 398
QY 611 AAGGGGTCTGTTAAGGAAGTGTGACAGGTACACAGCTGTGGTAAAGATGACTGCTCG 670
Db 397 AAGGGGTCTGTTAAGGAAGTGTGACAGGTACACAGCTGTGGTAAAGATGACTGCTCG 338
QY 671 GAGACAGGCTGAAACTGGACCAAGCTCATTTAGAGACAGTCAATTCGGCCCCGGGAAAA 730
Db 337 GAGACAGGCTGAAACTGGACCAAGCTCATTTAGAGACAGTCAATTCGGCCCCGGGAAAA 278
QY 731 GGGTTCTAGTTTTTAATGGAGGCTACAGAGGAATGAAGGCATCTCGAATCCATCAATG 790
Db 277 GGGTTCTAGTTTTTAATGGAGGCTACAGAGGAATGAAGGCATCTCGAATCCATCAATG 218
QY 791 AGAAGGCTTTTTCAGCCACGATAGTCAATGAAACTGGACCTTTGAAAGAGCGCAGAGTTG 850
Db 217 AGAAGGCTTTTTCAGCCACGATAGTCAATGAAACTGGACCTTTGAAAGAGCGCAGAGTTG 158
QY 851 AAGGTATTCAATATGAAGACATATCTAAACTTCCTTGAGTTTGAAATTTGATAACAACA 910
Db 157 AAGGTATTCAATATGAAGACATATCTAAACTTCCTTGAGTTTGAAATTTGATAACAACA 98
QY 911 CATTGAACTGTGAAGCATCAATTTGGTGTAGCCAGGCACTGTGTAACCTCTACTGTGT 970
Db 97 CATTGAACTGTGAAGCATCAATTTGGTGTAGCCAGGCACTGTGTAACCTCTACTGTGT 38
QY 971 TAGGGGATTTGTTTGTATTAAAAAATAAATAA 1003
Db 37 TAGGGGATTTCTTTGTTATTAAAAAATAAATAA 5
```

RESULT 15

BB749127 447 bp mRNA linear EST 16-OCT-2001
LOCUS BB749127 RIKEN full-length enriched, pooled tissues, pituitary,
DEFINITION etc. Mus musculus cDNA clone G030102N09 3', mRNA sequence.
ACCESSION BB749127
VERSION BB749127.1 GI:16153363
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidae; Muridae; Murinae; Mus.

REFERENCE
AUTHORS

1 (bases 1 to 447)
Akinura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K.,
Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
Matsumura,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akashira,S.,
Tanaka,Y., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akinura,T., et al.
2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gs.riken.jp, URL:http://genome.gsc.riken.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.

JOURNAL
COMMENT

Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gs.riken.jp, URL:http://genome.gsc.riken.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.

FEATURES
source

1..447
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="G030102N09"
/clone_lib="RIKEN full-length enriched, pooled tissues,
pituitary, etc."
/note="pooled tissues ; (tissue_type=pituitary gland,
dev_stage=adult, sex=male), (tissue_type=thymus,
dev_stage=adult, sex=male), (tissue_type=forelimb,
dev_stage=13 days embryo, sex=mixed), (tissue_type=testis,
dev_stage=13 days embryo, sex=male), (tissue_type=olfactory brain, dev_stage=adult, sex=male)"
(tissue_type=olfactory brain, dev_stage=adult, sex=male)"

ORIGIN

Query Match 39.3%; Score 433.4; DB 2; Length 447;
Best Local Similarity 99.6%; Pred. No. 8.5e-83;
Matches 445; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 624 GGAAGTGTACAGGTACACAGCTGTGGTAAAGATGACTCTCTGGAGACAGGCTGAA 683
DB 1 GGAAGTGTATGAAGGTACACAGCTGTGGTAAAGATGACTCTCTGGAGACAGGCTGAA 60
QY 684 ACTGGACAGACTCATTATAGACAGCTCCTCGGCCCGGGGAAAGGGTTCTAGTTT 743
DB 61 ACTGGACAGACTCATTATAGACAGCTCATTCCGGCCCGGGGAAAGGGTTCTAGTTT 120
QY 744 AAATGGAGGCTACAGAGGAATGAAGGCACTCTCGAATCCATCAATGAGAAGGCTTTTC 803
DB 121 AAATGGAGGCTACAGAGGAATGAAGGCACTCTCGAATCCATCAATGAGAAGGCTTTTC 180

QY 804 AGCCACGATAGTCATTGAAACTGGACCTTTGAAAGGACGACAGGTTGAAGGTATTCAATA 863
DB 181 AGCCACGATAGTCATTGAAACTGGACCTTTGAAAGGACGACAGGTTGAAGGTATTCAATA 240
QY 864 TGAAGACATATCTAAACTTGTCTTGAGTTTGAATAATTTGATAACACACACATTGAAACTGTG 923
DB 241 TGAAGACATATCTAAACTTGTCTTGAGTTTGAATAATTTGATAACACACACATTGAAACTGTG 300
QY 924 AAGCATCAAAATTTGGTGTAGCAAGGCACTGTGTAACCTCTACTGTGTTAGGGGATTTGTT 983
DB 301 AAGCATCAAAATTTGGTGTAGCAAGGCACTGTGTAACCTCTACTGTGTTAGGGGATTTGTT 360
QY 984 TTGTATTAAAAAATAAATAATCAATCTATTAAATACTAGTGAATAGTTGGGTAAATTTAT 1043
DB 361 TTGTATT-AAAAAATAAATAATCAATCTATTAAATACTAGTGAATAGTTGGGTAAATTTAT 419
QY 1044 AATAAAATCTATGTTTTTTTTTTTAAAGTGT 1070
DB 420 AATAAAATCTATGTTTTTTTTTTTAAAGTGT 446

Search completed: November 27, 2005, 00:57:45
Job time : 6166.01 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:00:02 ; Search time 228.214 Seconds
(without alignments)
8583.479 Million cell updates/sec

Title: US-09-555-529-2
Perfect score: 1102
Sequence: 1 atggcgaagtcggatttct.....aaaaaaaaaaaaaaaaaaaa 1102

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2/6/ptodata/1/ina/1/COMB.seq: *
2: /cgn2/6/ptodata/1/ina/5/COMB.seq: *
3: /cgn2/6/ptodata/1/ina/6A/COMB.seq: *
4: /cgn2/6/ptodata/1/ina/6B/COMB.seq: *
5: /cgn2/6/ptodata/1/ina/H/COMB.seq: *
6: /cgn2/6/ptodata/1/ina/PTUS/COMB.seq: *
7: /cgn2/6/ptodata/1/ina/PP/COMB.seq: *
8: /cgn2/6/ptodata/1/ina/RE/COMB.seq: *
9: /cgn2/6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	118.8	10.8	538	3	US-09-270-767-379, App
c 2	118.8	10.8	538	3	US-09-270-767-15661, A
c 3	90	8.2	23645	3	US-09-949-016-13916, A
c 4	69.4	6.3	7218	2	US-08-232-463-14, App1
5	61.8	5.6	468	3	US-09-270-767-9580, Ap
6	61.8	5.6	468	3	US-09-270-767-2482, A
7	60	5.4	825	3	US-09-248-796A-5938, Ap
8	59	5.4	2229	3	US-08-207-954-6, App1
9	59	5.4	2510	2	US-08-888-982A-42, App1
10	59	5.4	2510	3	US-09-462-261-42, App1
11	59	5.4	2510	3	US-09-506-073-89, App1
12	59	5.4	2510	3	US-10-057-550C-67, App1
13	54.8	5.0	3680	3	US-09-647-390-15, App1
14	54.6	5.0	332	3	US-09-621-976-16031, A
15	53.2	4.8	1752	3	US-09-292-225-14, App1
c 16	53.2	4.8	1752	3	US-09-292-225-16, App1
17	53.2	4.8	1753	3	US-09-149-476-56, App1
18	52.8	4.8	1039	3	US-09-902-540-1280, Ap
19	52.8	4.8	2797	3	US-09-482-273-74, App1
20	52.4	4.8	247299	3	US-09-949-016-17590, A
21	52.2	4.7	1553	3	US-09-280-116-10, App1
22	51.6	4.7	1976	3	US-09-920-759-10, App1
23	51.2	4.6	785	3	US-09-205-258-73, App1
24	51.2	4.6	785	3	US-10-004-860-73, App1

25	51.2	4.6	41736	3	US-09-949-016-17091, A
26	51	4.6	306	3	US-09-621-976-16035, A
27	51	4.6	326	3	US-09-621-976-16024, A
28	51	4.6	329	3	US-09-621-976-16012, A
29	51	4.6	332	3	US-09-621-976-16053, A
30	51	4.6	333	3	US-09-621-976-16032, A
31	51	4.6	333	3	US-09-621-976-16045, A
32	51	4.6	334	3	US-09-621-976-16044, A
33	51	4.6	336	3	US-09-621-976-16013, A
34	51	4.6	347	3	US-09-621-976-16026, A
35	51	4.6	357	3	US-09-621-976-16058, A
36	51	4.6	359	3	US-09-621-976-16008, A
37	51	4.6	359	3	US-09-621-976-16019, A
38	51	4.6	362	3	US-09-621-976-16010, A
39	51	4.6	365	3	US-09-621-976-16042, A
40	51	4.6	371	3	US-09-621-976-16048, A
41	50.6	4.6	305	3	US-09-621-976-16020, A
42	50.6	4.6	306	3	US-09-621-976-16057, A
43	50.6	4.6	327	3	US-09-621-976-16018, A
44	50.6	4.6	387	3	US-10-125-258-104, App
45	50.2	4.6	332	3	US-09-621-976-16050, A

ALIGNMENTS

RESULT 1

US-09-270-767-379/c
; Sequence 379, Application US/09270767
; Patent No. 6703491

; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 379
; LENGTH: 538
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-379

Query Match 10.8%; Score 118.8; DB 3; Length 538;
Best Local Similarity 56.3%; Pred. No. 1.6e-18;
Matches 222; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

Qy	496	ATCATGGAGCTCGAAGAGGAAAAAGGACCGCACGACAGCGCTGGTTACAGCCG	555
Db	522	ATCAGCAGGAGGAAACCAAAAGGAGCGTGCCACCGCAAGACTACTGGCTGCAAG	463
Qy	556	GGGATCGTTGTGAAAAATTATAACGAGAGAGCTTTGGGAGAAATATCAAGAGAAAGGG	615
Db	462	GGTATCGTGTCAAAATTTATTTCCAAATCCATCGGGGAAAAGTTCTTCAACAAAAAGCG	403
Qy	616	GTCTTAAAGAGAGTATTGACAGGTACACAGCTGTGGTAAAGATGACTCTGGAGAC	675
Db	402	GTGTCTCGGACGTAATTGACATATCAGGGGAAAATCAAGTTCTTTGGAGACTGGGGAA	343
Qy	676	AGCTGAAACTCGACACAGACTCATTTTAGACACAGTCAATCCGGCCCGGGGAAAAGGTT	735
Db	342	AAGCTAAAGTGAATCAAGTCAATTTGGACGCGTAAATCCCGCTTTGGACAAAGCCTGTC	283
Qy	736	CTAGTTTTTAATGGAGCTACAGAGGAAATGAAGGCACTCTCGAATCCATCAATGAGAG	795
Db	282	ATGTTGTTAATGGCGTTATCGGGATCGGAGCTCTGCTAAGGAAACTGGACGAGCG	223
Qy	796	CTTTTTCAGCCAGTAGTCAATTTGAACCTGGACCTTTGAAGGACCGCAGAGTTGAAGGT	855
Db	222	AGATATTTCAGTCAGCGTGGAAATATTGACCGGCTCTCAAGGCGAGATTTGTAGCAAC	163
Qy	856	ATTCATATGAGACATATCTAAACTTGTGTTAG	889

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13916
; LENGTH: 23645
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13916

Query Match      8.2%; Score 90; DB 3; Length 23645;
Best Local Similarity 86.8%; Pred. No. 3.5e-11;
Matches 99; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 ATGGCCAGAGTCGGATTTCTGAGCCCAAGCCCAATGCGCAATAGTAATTAAGTCCAAAGGG 60
DB 1763 ATGGGGAAGTCGGATTTCTTACTCCCAAGGCTATCGCCCAACAGGATCAAGTCCCAAGGG 1704
QY 61 CTCACAGAGCTTCGCTGGTACTGCCAGATGTGCCAAAAGCAATGCCGCGACGAG 114
DB 1703 CTGCAGAGCTACGCTGGTATTGCCAGATGTGCCAGAGCAGTGCCTGGGACGAG 1650

RESULT 4
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935.313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match      6.3%; Score 69.4; DB 2; Length 7218;
Best Local Similarity 2.6%; Pred. No. 1.7e-06;
Matches 10; Conservative 236; Mismatches 137; Indels 0; Gaps 0;
```

```
Db 162 GTGCAATACGAGATATATCTAAACTACATGGCG 129

RESULT 2
US-09-270-767-15661/c
; Sequence 15661, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 15661
; LENGTH: 538
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-15661

Query Match      10.8%; Score 118.8; DB 3; Length 538;
Best Local Similarity 56.3%; Pred. No. 1.6e-18;
Matches 222; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

QY 496 ATCATGGAGCTCGAAGAGGAAAGAAAGCGCACGAGCAGCCCTGGTTACAGCGG 555
DB 522 ATCAAGCAGGAGGAAAGCAAGAGGAGCGTGCCCAACCGCAAGACTACTGGCTGCACAAG 463
QY 556 GGGATCGTTGTGAAATTTATACGAGAAGCTTGGGGAGAAATATACAAAGAAAGGG 615
DB 462 GGTATCGTGTCAAATTTATTTCCAAATCCATGGCGGAAAGTTCTTCAACAAAGGCG 403
QY 616 GTCGTTAAGCAGAGTATGACAGTACACAGCTGTGTTAAAGATGACTGACTCTGGAGAC 675
DB 402 GTTGCTCGACGTAATTGACAGATATCAGGCGGAAATCAAGTTCTTGAGACTGGGGAA 343
QY 676 AGCTGAAACTGGACACAGACTCATTTAGAGACAGTCAATTCGCGCCCGGGGAAAGGGTT 735
DB 342 AAGCTAAAAGTGGATCAAGCTCATTTGGAGACGGTAATCCCGCTTTGGACAAGCCTGTC 283
QY 736 CTAGTTTTAAATGGAGGCTACAGAGAAATGAAGGCACTCTGAAATCCATCATGAGAAG 795
DB 282 ATGCTGGTTAATGCGCTTATCGGGATCGAGGCTCTGTAAGGAAACTGGACGAGCGC 223
QY 796 GCTTTTTCAGCCACGATAGTCATTGAACTTGAAGACCTTTGAAAGGACGACAGAGTTGAAGT 855
DB 222 AGATATTCAGTCAGCTGGAATATTGACGGTCTCTCAAGGCGAATTTGTAGACAAC 163
QY 856 ATTCATATGAAGACATATCTAAACTTGCTTGAG 889
DB 162 GTGCAATACGAGATATATCTAAACTACATGGCG 129

RESULT 3
US-09-949-016-13916/c
; Sequence 13916, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
```

```
QY 395 GTCCACTGAAGCTGCTGGGAGCGCAGCATCCGGGAAACGGAAGAGCTTTCACAGAGCT 454
Db 1438 GTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1379
QY 455 CCCCCAGCCTCGAAGAAGAAAGCTCGGCCCTGGATGAGATCATGGAGCTCGAAGAGG 514
Db 1378 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1319
QY 515 AAAAGAAAAGGACCGCAGCAGACAGCGCTGTTACAGCGGGGATCGTGTGGAATAATTA 574
Db 1318 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1259
QY 575 TAACGAAGAAGCTGGGGAGAAATATCACAGAAGAAAGGGTCGTTAAGGAAGTCATTG 634
Db 1258 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1199
QY 635 ACAGGTACACAGCTGTGGTAAAGATGACTGACTCTGGAGACAGGCTGAAACTGGACGAG 694
Db 1198 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1139
QY 695 CTCATTTAGACAGCTATCCGGCCCGGGGAAAGGGTTCAGTTTAAATGGAGGCT 754
Db 1138 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1079
QY 755 ACAGGGAATGAAGCACTCTC 777
Db 1078 RRRRRRRRRRRATCGCAAGCTC 1056
```

```
RESULT 5
US-09-270-767-9580
; Sequence 9580, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9580
; LENGTH: 468
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-9580
```

```
Query Match 5.6%; Score 61.8; DB 3; Length 468;
Best Local Similarity 64.1%; Pred. No. 4.4e-05;
Matches 93; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 958 AACTCTACTGTGTAGGGGATTTGTTTGTATTAATAAAGAAATCATCTATTTAAAT 1017
Db 219 AGCTTTACGCTTTTATGAGGATTCATATATTAAGAAAGAAATTTATGCTTTAAT 278
QY 1018 ACTAGTGAATGTTGGTGAATTTATTAATAAATCTATGTTTTTTTAAAGTGTAAAAA 1077
Db 279 TTATATAAAATTTTAAATAATTCATAATCAATATATATGTTGTTAGTTTCATGAACA 338
QY 1078 AAAAAAAAAAAAAAAAAAAAAA 1102
Db 339 CAATGAAAAAAAAAAAAACAGAAAAA 363
```

```
RESULT 6
US-09-270-767-24862
; Sequence 24862, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
```

```
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24862
; LENGTH: 468
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-24862

Query Match 5.6%; Score 61.8; DB 3; Length 468;
Best Local Similarity 64.1%; Pred. No. 4.4e-05;
Matches 93; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 958 AACTCTACTGTGTAGGGGATTTGTTTGTATTAATAAAGAAATCATCTATTTAAAT 1017
Db 219 AGCTTTACGCTTTTATGAGGATTCATATATTAAGAAAGAAATTTATGCTTTAAT 278
QY 1018 ACTAGTGAATGTTGGTGAATTTATTAATAAATCTATGTTTTTTTAAAGTGTAAAAA 1077
Db 279 TTATATAAAATTTTAAATAATTCATAATCAATATATATGTTGTTAGTTTCATGAACA 338
QY 1078 AAAAAAAAAAAAAAAAAAAAAA 1102
Db 339 CAATGAAAAAAAAAAAAACAGAAAAA 363
```

```
RESULT 7
US-09-248-796A-5938
; Sequence 5938, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 5938
; LENGTH: 825
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-5938
```

```
Query Match 5.4%; Score 60; DB 3; Length 825;
Best Local Similarity 49.0%; Pred. No. 0.00014;
Matches 193; Conservative 0; Mismatches 195; Indels 6; Gaps 1;

QY 1 ATGGCAAGTCGGAATTTCTGAGCCCAAGGCCATCGCCAATAGAATTAAGTCCAAAGGG 60
Db 106 ATGGCAAGGCGAGAGTTTGGAAACAGCAAAATATCAATCTAAAAAGCTACGAGCGGCGA 165
QY 61 CTCAGAGGCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
Db 166 TTACAAAAATTTGAAGTTCTTATTCGCAATTTGTTGTTTCTTAAACAGTGTGAGATTC 225
QY 121 TTTAAGTGTCTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 174
Db 226 TTTAAAAACCATCTATCTTCCCGCTGCACATTAAGAGGTTCCGAAATACATGATCT 285
QY 175 GAAAAACCTCAGCAGTTTATGGAATTTATTTTCAGAGGAATTCGGAATGATCTTCTGG 234
Db 286 GGTGATAGCTCCAAAGCTTAATAGAAACCTACTCTACAAAATTTCCAAGATAAGTTTAT 345
QY 235 CTTCTGAGGCGAGCGCTTTCGCACTTAAAGGGTCCACAAACCAATGCTCTCAATGAAT 294
Db 346 TTGCTCCGAATCAACCAATGGAAGTAAATTTATCAATGCCCAATAGTTTCTACCAAG 405
QY 295 ATCAGCCACCGAGAGACATCCACATGAACGCTACCCAGTGGGAGACACTGACCGACT 354
```

```
Db 406 ATCCGGAGGAGATCATATCCATATGAATCCACCAGATGGAGAAAGTCTAACTTCATTT 465
QY 355 ACCAAGTGGCTGGCGAGAGAGGCTTGTGTAAAG 388
Db 466 ATAARACACTTGGGTAATAAATGGATGTGCAAG 499

RESULT 8
US-08-207-954-6
; Sequence 6, Application US/08207954
; Patent No. 6689560
; GENERAL INFORMATION:
; APPLICANT: RAPP, ULF
; APPLICANT: APP, HARALD
; APPLICANT: STORM, STEPHEN M.
; TITLE OF INVENTION: RAF PROTEIN KINASE THERAPEUTICS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: 1100 NEW YORK AVE., N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/207,954
; FILING DATE: 18-MAR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/749,931
; FILING DATE: 23-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T.
; REGISTRATION NUMBER: 26,581
; REFERENCE/DOCKET NUMBER: 5683/82731
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3067
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 229 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-207-954-6
Query Match 5.4%; Score 59; DB 3; Length 2229;
Best Local Similarity 69.6%; Pred. No. 0.00033;
Matches 80; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 988 ATTAATAAAAAAAAAAATCATCTATTAAATAGTGAATAGTTGGGTAAATTTTATAATA 1047
Db 2105 AGTAGCAACAAGGAATAAATGAACATATGTTGCTTATATGTTAAATTTGAATAAAA 2164
QY 1048 AAATCTATCTTTTTTTTAAAGTGTAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1102
Db 2165 TACTCTCTTTTTTTTTTAAAGTGGAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2219

RESULT 9
US-08-888-982A-42
; Sequence 42, Application US/08888982A
; Patent No. 5981731
; GENERAL INFORMATION:
; APPLICANT: Brett P. Monia
```

```
; TITLE OF INVENTION: Antisense Oligonucleotide Modulation
; TITLE OF INVENTION: of raf Gene Expression
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jane Massey Licata, Esq.
; STREET: 66 East Main Street
; CITY: Marlton
; STATE: NJ
; COUNTRY: USA
; ZIP: 08053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
; COMPUTER: IBM 486
; OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/888,982A
; FILING DATE: Herewith
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/756,806
; FILING DATE: No. 5981731ember 26, 1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07111
; FILING DATE: May 31, 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/250,856
; FILING DATE: May 31, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Jane Massey Licata
; REGISTRATION NUMBER: 32,257
; REFERENCE/DOCKET NUMBER: ISPH-0212
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (609) 779-2400
; TELEFAX: (609) 779-8488
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2510
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; ANTI-SENSE: NO
US-08-888-982A-42
Query Match 5.4%; Score 59; DB 2; Length 2510;
Best Local Similarity 69.6%; Pred. No. 0.00035;
Matches 80; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 988 ATTAATAAAAAAAAAAATCATCTATTAAATAGTGAATAGTTGGGTAAATTTTATAATA 1047
Db 2386 AGTAGCAACAAGGAATAAATGAACATATGTTGCTTATATGTTAAATTTGAATAAAA 2445
QY 1048 AAATCTATCTTTTTTTTAAAGTGTAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1102
Db 2446 TACTCTCTTTTTTTTTTAAAGTGGAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2500

RESULT 10
US-09-462-261-42
; Sequence 42, Application US/09462261
; Patent No. 6391636
; GENERAL INFORMATION:
; APPLICANT: Brett P. Monia
; TITLE OF INVENTION: Antisense Oligonucleotide
; Modulation of raf Gene Expression
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jane Massey Licata, Esq.
; STREET: 66 East Main Street
; CITY: Marlton
; STATE: NJ
; COUNTRY: USA
; ZIP: 08053
```

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
; COMPUTER: Pentium
; OPERATING SYSTEM: Windows 95
; SOFTWARE: WORDPERFECT 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/462,261
; FILING DATE: 01-Mar-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/756,806
; FILING DATE: No. 6391636ember 26, 1996
; APPLICATION NUMBER: PCT/US95/07111
; FILING DATE: May 31, 1995
; APPLICATION NUMBER: 08/250,856
; FILING DATE: May 31, 1994
; APPLICATION NUMBER: 08/888,982
; FILING DATE: July 7, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Jane Massey Licata
; REGISTRATION NUMBER: 32,257
; REFERENCE/DOCKET NUMBER: ISPH-0312
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (609) 779-2400
; TELEFAX: (609) 810-1454
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2510
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-09-462-261-42

Query Match      5.4%; Score 59; DB 3; Length 2510;
Best Local Similarity 69.6%; Pred. No. 0.00035;
Matches 80; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 988 ATTAAAAAATAATCATCTATTAAATAGTGAATAGTTGGTAAATTTATAATA 1047
DB 2386 AGTAGCAACAAAGGAAATAAATGAACATATGTTGCTTATATGTTAAATTTGAATAA 2445
QY 1048 AAATCTATGTTTTTTTTTAAAGTGTAAAAAATAAAAAAAAAAAAAAAAAAAAA 1102
DB 2446 TACTCTCTTTTTTTTTTAAAGTGTGAAAAAATAAAAAAAAAAAAAAAAAAAAA 2500

RESULT 12
US-10-057-550C-67
; Sequence 67, Application US/10057550C
; Patent No. 6806258
; GENERAL INFORMATION:
; APPLICANT: Monia, Brett P
; TITLE OF INVENTION: Antisense Oligonucleotide Modulation of raf Gene Expression
; FILE REFERENCE: ISPH-0625
; CURRENT APPLICATION NUMBER: US/10/057,550C
; CURRENT FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: US 09/506,073
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: US 09/143,214
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: US 08/756,806
; PRIOR FILING DATE: 1996-11-26
; PRIOR APPLICATION NUMBER: PCT/US95/07111
; PRIOR FILING DATE: 1995-05-31
; PRIOR APPLICATION NUMBER: US 08/250,856
; PRIOR FILING DATE: 1994-05-31
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 67
; LENGTH: 2510
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-057-550C-67

Query Match      5.4%; Score 59; DB 3; Length 2510;
Best Local Similarity 69.6%; Pred. No. 0.00035;
Matches 80; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 988 ATTAAAAAATAATCATCTATTAAATAGTGAATAGTTGGTAAATTTATAATA 1047
DB 2386 AGTAGCAACAAAGGAAATAAATGAACATATGTTGCTTATATGTTAAATTTGAATAA 2445
QY 1048 AAATCTATGTTTTTTTTTAAAGTGTAAAAAATAAAAAAAAAAAAAAAAAAAAA 1102
DB 2446 TACTCTCTTTTTTTTTTAAAGTGTGAAAAAATAAAAAAAAAAAAAAAAAAAAA 2500

RESULT 13
US-09-647-390-15
; Sequence 15, Application US/09647390
; Patent No. 6465636
; GENERAL INFORMATION:
; APPLICANT: Stuijver, Maarten
; APPLICANT: Cueters, Jerome
; APPLICANT: Simons, Lambertus
; TITLE OF INVENTION: Pathogen-Inducible Promoter
; FILE REFERENCE: MOG 57707/UST
; CURRENT APPLICATION NUMBER: US/09/647,390
; CURRENT FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: EP 98201024.1
; PRIOR FILING DATE: 1998-04-01
```

```
; PRIOR APPLICATION NUMBER: PCT/EP99/02178
; PRIOR FILING DATE: 1999-03-25
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 3680
; TYPE: DNA
; ORGANISM: Helianthus annuus
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1889)
; NAME/KEY: CDS
; LOCATION: (1890)..(3503)
US-09-647-390-15

Query Match          5.0%; Score 54.8; DB 3; Length 3680;
Best Local Similarity 54.5%; Pred. No. 0.0039;
Matches 110; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 895 AAATTGTGATAACACACATGAAACCTGTCGAGCATCAAAATGGTGTAGCCAGGCACGTG 954
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 697 AATTTTTTTTAAATAAATGAAATTTTAAAGAGATCATTTTCTAAAGATCCGTAGCG 756
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 955 TGTAACTCTACTGCTGTAGGGATTGTTTGTATTAAAAAATAAATCATCTATTTA 1014
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 757 AGTAAAGTTATGATGTTGCTACCTTTTATGTTCTTATTCATCTGTTTAAATA 816
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1015 AATACTAGTGAATGAGTGGGTAAATTTAATAATAAATCTATGTTTTTTTAAAGTGTA 1074
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 817 TAAAGAGATAAGGAGTGTGTAAACAAATATAAGAGTGGTTAAAGGTAAAAA 876
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1075 AAAAAAAAAAAAAAAAAAAAAA 1096
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 877 AAAAAAAAAAAAAAAAAAAAAA 898
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
US-09-621-976-16031
; Sequence 16031, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S. Y.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 16031
; LENGTH: 332
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-16031

Query Match          5.0%; Score 54.6; DB 3; Length 332;
Best Local Similarity 65.2%; Pred. No. 0.002;
Matches 75; Conservative 3; Mismatches 37; Indels 0; Gaps 0;

QY 988 ATTAAAAAATAAATCATCTATTAAATACACTAGTGAATAGTTGGTAAATTTATAATA 1047
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 192 ATGAAATAGAAAAAATAATTTTCGATAGAAAAATAAATAGAAAAATTTAAAAA 251
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1048 AAATCTATGTTTTTTTTTAAGTGTAATAAAAAAAAAAAAAAAAAAAAAA 1102
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 252 CAACCCAGCCTCATCTATGAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 306
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
US-09-292-225-14
; Sequence 14, Application US/09292225
; Patent No. 6455686
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/09/292,225
; CURRENT FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: 60/098,909
; EARLIER FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/085,295
; EARLIER FILING DATE: 1998-05-13
; EARLIER APPLICATION NUMBER: 60/098,565
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: 09/062,013
; EARLIER FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 1752
; TYPE: DNA
; ORGANISM: Dermatophagoides farinae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1665)
US-09-292-225-14

Query Match          4.8%; Score 53.2; DB 3; Length 1752;
Best Local Similarity 61.6%; Pred. No. 0.0072;
Matches 85; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 965 CTGTGTTAGGGGATTTGTTTGTATTAAAAAATAAATCATCTATTAAATAGTAGTG 1024
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1614 CTGTCCACCGGCACACTATTGTTGTCAAGAAAAATTGACTTGTATAGGCGAATAATCTG 1673
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1025 AATAGTTGGGTAAATTTTATAAATAAATCTATGTTTTTTTTTAAGTGTAATAAATAA 1084
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1674 AAAAAAATAATCAATTAATAAATTTAAATAATTCATTTTTTAATATGAAAAAATTC 1733
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1085 AAAAAAAAAAAAAAAAAAAAAA 1102
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1734 AAAAAAAAAAAAAAAAAAAAAA 1751
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Search completed: November 27, 2005, 01:10:38
Job time : 230.214 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:00:41 ; Search time 1270.42 Seconds
(without alignments)
7173.088 Million cell updates/sec

Title: US-09-555-529-2
Perfect score: 1102
Sequence: 1 atggcgcaagtcggattttctt.....aaaaaaaaaaaaaaaaaaaaa 1102

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA.Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	309	28.0	461	7	US-10-242-535A-29177 Sequence 29177, A
2	309	28.0	461	7	US-10-085-783A-29177 Sequence 29177, A
3	269.2	24.4	591	5	US-10-106-698-1187 Sequence 1187, Ap
4	172.4	15.6	538	7	US-10-437-963-89433 Sequence 89433, A
5	169.2	15.4	2046	8	US-10-425-115-157125 Sequence 157125, A
6	168.8	15.3	1614	7	US-10-425-114-2975 Sequence 2975, Ap
7	168.8	15.3	1713	8	US-10-425-114-22614 Sequence 22614, A
8	167.6	15.2	1713	8	US-10-739-930-2955 Sequence 2955, Ap
9	165	15.0	777	7	US-10-424-599-25484 Sequence 25484, A
10	163	14.8	1241	10	US-11-097-143-41039 Sequence 41039, A
11	163	14.8	3313	10	US-11-097-143-41038 Sequence 41038, A
12	163	14.8	4394	10	US-11-097-143-41017 Sequence 41017, A
13	163	14.8	4582	10	US-11-097-143-24532 Sequence 24532, A
14	150.4	13.6	584	7	US-10-767-701-25587 Sequence 25587, A
15	130.8	11.9	549	3	US-09-991-936-1424 Sequence 1424, Ap
16	130.8	11.9	549	3	US-10-978-245-1424 Sequence 1424, Ap
17	95.4	8.7	300	9	US-10-779-543-7646 Sequence 7646, Ap
18	93.6	8.5	270	3	US-09-294-0938-3089 Sequence 3089, A
19	91.2	8.3	739	7	US-10-424-599-53324 Sequence 53324, A
20	81.2	7.4	516	4	US-09-925-065A-549989 Sequence 549989, A
21	67.2	6.1	521	8	US-10-425-115-116286 Sequence 116286, A
22	66.2	6.0	375	8	US-10-357-930-44930 Sequence 44930, A
23	65	5.9	65	3	US-09-908-975-24642 Sequence 24642, A

ALIGNMENTS

RESULT 1

US-10-242-535A-29177
; Sequence 29177, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liaw, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29177
; LENGTH: 461
; TYPE: DNA
; ORGANISM: Human
US-10-242-535A-29177

Query Match 28.0%; Score 309; DB 7; Length 461;
Best Local Similarity 90.5%; Pred. No. 1.6e-56;
Matches 352; Conservative 0; Mismatches 35; Indels 2; Gaps 2;

Qy	1	ATCGGCAAGTCGGAATTTCTGAGCCGCCAAGCCCATGCCCAATAGTAATTAAGTCCAAAGGG	60
Db	8	ATGGGGAAGTCGGATTTTCTTACTCCCAA-GCTATGCCCAACAGGATCAAGTCCAGGGG	66
Qy	61	CTCAGAGGTTTCCTGGTACTGCCAGATGTCGCAAGCAATGCGCGCAGAGAAATGGC	120
Db	67	CTGCAAGGCTACGCTGGTATTCAGATGTCGCAAGCAGATGCCCGGAGAGAAATGGC	126
Qy	121	TTTAAGTGTCACGTGTATGCTGAATCTCATCAAGACAACCTGTTGCTGGCTTCAGAAAAC	180
Db	127	TTTAAGTGTCATGTATGTCGGAATCTCACCAGAGACAATATTGCTGGCTTCAGAAAAT	186
Qy	181	CCTCAGAGTTATGGAATATTTTTCAGAGAAATTCGAGAAATGACTTTCTGGAACTTCTG	240
Db	187	CCTCAGAGTTATGGAATATTTTTCAGAGAAATTCGAGAAATGACTTTCTAGAACTTCTC	246

;; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

;; FILE REFERENCE: 38-21(53221)B

;; CURRENT APPLICATION NUMBER: US/10/437,963

;; NUMBER OF SEQ ID NOS: 204966

;; CURRENT FILING DATE: 2003-05-14

;; SEQ ID NO 89433

;; LENGTH: 538

;; TYPE: DNA

;; ORGANISM: Oryza sativa

;; FEATURE:

;; OTHER INFORMATION: Clone ID: PAT_MRT4530_88190C.1

US-10-437-963-89433

Query Match 15.6%; Score 172.4; DB 7; Length 538;

Best Local Similarity 65.1%; Pred. No. 4.8e-27;

Matches 254; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 1 ATGGGCAAGTCGGATTCTCTGAGCCCAAGGCCATCGCCAAATAGAATTAAATCCAAAGGG 60

DB 124 ATGGGGAAGCACGAGTCTCTGAGCCGAGGCCGATCGGAAAGGATCAAGGCGAAGGGG 183

QY 61 CTCGAGAGCTTCGCTGGTACTGCCAGATGTGCCAAAGCAATGCCGCGACGAGATGGC 120

DB 184 CTCGAGAGCTTCGCTGGTACTGCCAGATGTGCCAGAGCAGTGGCCGACGAGATGGC 243

QY 121 TTAAAGTGCACGTATGTCTGAATCTCATCAAGACAACTGTGCTGGCTTCAGAAAC 180

DB 244 TTCAAGTGCCACGTATGTCTGAATCTCATCAAGACAACTGTGCTGGCTTCAGAAAC 303

QY 181 CCTCAGCAGTTTATGATTTATTTTTCAGAGGAATTCGGAATGACATTTCTGGAATCTTG 240

DB 304 CCGACCGAGTCTCGAGGGCTTCTCGAGGAGTTCCTCGACGCTTCCTCACCTTGCTC 363

QY 241 AGCGGAGCTTGGCACTAAAGGCTCCAAACAACTTCTTACAAATGAATACATCAGC 300

DB 364 CGCGGGCCACCGACACTCCCGATCGCGCCACCGTCTCTACACGAGTTCAATCGCC 423

QY 301 CACCGAGAGCAATCCACATGAAACGCTACCCAGTGGGAGACACTGACCGACTTTACCAAG 360

DB 424 GACCGCCACCACTCCACATGAATCCACGCTGGGGCCACGCTCACCGAGTTTGTCAAG 483

QY 361 TGCGTGGGCGAGAGGGCTTGTGTAAGGT 390

DB 484 TTCTCGGGCGGAGGGCCACTGCAAGTT 513

RESULT 5

US-10-425-115-157125

;; Sequence 157125, Application US/10425115

;; Publication No. US20040214272A1

;; GENERAL INFORMATION:

;; APPLICANT: La Rosa, Thomas J.

;; APPLICANT: Kovalic, David K.

;; APPLICANT: Zhou, Yihua

;; APPLICANT: Cao, Yongwei

;; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

;; TITLE OF INVENTION: Plants

;; FILE REFERENCE: 38-21(53222)B

;; CURRENT APPLICATION NUMBER: US/10/425,115

;; CURRENT FILING DATE: 2003-04-28

;; NUMBER OF SEQ ID NOS: 369326

;; SEQ ID NO 157125

;; LENGTH: 2046

;; TYPE: DNA

;; ORGANISM: Zea mays

;; FEATURE:

;; OTHER INFORMATION: Clone ID: MRT4577_74879C.1

US-10-425-115-157125

Query Match 15.4%; Score 169.2; DB 8; Length 2046;

Best Local Similarity 64.6%; Pred. No. 4.3e-26;

Matches 252; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

QY 1 ATGGGCAAGTCGGATTCTCTGAGCCCAAGGCCATCGCCAAATAGAATTAAATCCAAAGGG 60

DB 298 ATGGGGAAGCACGAGTCTCTGAGCCGAGGCGATCGGAAATCGGATCAAGGCGAAGGGG 357

QY 61 CTCGAGAGCTTCGCTGGTACTGCCAGATGTGCCAAAGCAATGCCGCGACGAGATGGC 120

DB 358 CTGCAGAGCTTCGCTGGTACTGCCAGATGTGTCAAGAGCAGTGGCCGCGACGAGAACGGG 417

QY 121 TTAAAGTGCACGTATGTCTGAATCTCATCAAGACAACTGTGCTGGCTTCAGAAAC 180

DB 418 TTCAAGTGCCACGTATGTCTGAATCTCATCAAGACAACTGTGCTGGCTTCAGAAAC 477

QY 181 CCTCAGCAGTTTATGATTTATTTTTCAGAGGAATTCGGAATGACATTTCTGGAATCTTG 240

DB 478 CCGACCGGCTCTCGAGGGCTTCTCGAGGAAATCTCTCGAGTCTTCTCTCCCTCATC 537

QY 241 AGCGGAGCTTGGGCACTAAAGGCTCCAAACAACTTGTCTACAAATGAATACATCAGC 300

DB 538 CGCGGGCCACCGCCACTCCCGCGTGGCCGCGCCACCGTTGTCTACAAACGAGTACATCGCC 597

QY 301 CACCGAGAGCAATCCACATGAAACGCTACCCAGTGGGAGACACTGACCGACTTTTACCAAG 360

DB 598 GACCGGCAACGCTCCACATGAATCTACGCGTGGGCCACGCTCACCGAGTTCTGTCAG 657

QY 361 TGCGTGGGCGAGAGGGCTTGTGTAAGGT 390

DB 658 CTCCTGGGCGGAGGGTACTGCAAGTT 687

RESULT 6

US-10-425-114-2975

;; Sequence 2975, Application US/10425114

;; Publication No. US20040034888A1

;; GENERAL INFORMATION:

;; APPLICANT: Liu, Jingdong

;; APPLICANT: Zhou, Yihua

;; APPLICANT: Kovalic, David K.

;; APPLICANT: Screen, Steven E.

;; APPLICANT: Tabaska, Jack E.

;; APPLICANT: Cao, Yongwei

;; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

;; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

;; FILE REFERENCE: 38-21(53113)B

;; CURRENT APPLICATION NUMBER: US/10/425,114

;; CURRENT FILING DATE: 2003-04-28

;; NUMBER OF SEQ ID NOS: 73128

;; SEQ ID NO 2975

;; LENGTH: 1614

;; TYPE: DNA

;; ORGANISM: Zea mays

;; FEATURE:

;; OTHER INFORMATION: Clone ID: 700238628_FLI

US-10-425-114-2975

Query Match 15.3%; Score 168.8; DB 7; Length 1614;

Best Local Similarity 64.7%; Pred. No. 4.7e-26;

Matches 251; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 1 ATGGGCAAGTCGGATTCTCTGAGCCCAAGGCCATCGCCAAATAGAATTAAATCCAAAGGG 60

DB 80 ATGGGGAAGCACGAGTCTCTGAGCCGAGGCGATCGGAAATCGGATCAAGGCGAAGGGG 139

QY 61 CTCGAGAGCTTCGCTGGTACTGCCAGATGTGCCAAAGCAATGCCGCGACGAGATGGC 120

DB 140 CTGCAGAGCTTCGCTGGTACTGCCAGATGTGTCAAGAGCAGTGGCCGCGACGAGAACGGG 199

QY 121 TTAAAGTGCACGTATGTCTGAATCTCATCAAGACAACTGTGCTGGCTTCAGAAAC 180

DB 200 TTCAAGTGCCACGTATGTCTGAATCTCATCAAGAGGAGATGCGAGTGTTCGGCATGGCC 259

QY 181 CCTCAGCAGTTTATGATTTATTTTTCAGAGGAATTCGGAATGACATTTCTGGAATCTTG 240

```
Db 260 CCCGACCGCGTCGTCCAGGGCTTCTCCGAGGAGTTCTCGAGTCTTCTCCTCCTCATC 319
QY 241 AGCGCAGCGTTTGGCACTAAAGGFTCCACAACATTGTCTACAAATGAATACATCAGC 300
Db 320 CGCCGCGCGCACCGCACTCCGCGTCCGCGCACCGTGTCTACAAAGTACATCGCC 379
QY 301 CACCGAGACCATCCACATGAACGCTACCCAGTGGGAGACACTGACCGACTTTACCAAG 360
Db 380 GACCGGCACACGTCACATCACTCCAGCGGTGGGCCACGCTCACCGAGTTCGTCAAG 439
QY 361 TGGCTGGGCAGAGAGGGCTTGTGTAAG 388
Db 440 TTCCTGGGCGCGAGGGGTACTGTAAAG 467

RESULT 7
US-10-425-114-22614
; Sequence 22614, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 22614
; LENGTH: 1713
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3591-011-E3_FLI
US-10-425-114-22614
Query Match 15.3%; Score 168.8; DB 7; Length 1713;
Best Local Similarity 64.7%; Pred. No. 4.8e-26;
Matches 251; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 1 ATGGGCAAGTCGGATTCTTCGAGCCCAAGGCCATCGCCAATAGAAATTAAGTCCAAAGGG 60
Db 239 ATGGGCAAGTCGGATTCTTCGAGCCGGAAGCGATCGCAACCGGATCAAGGCGAAGGGG 298
QY 61 CTCGAGAGCTCGGTGTTACTGCGCAGATGTGCCAAAAGCAATGCCGCGACGAGATGGC 120
Db 299 CTCGAGAGCTCGGTGTTACTGTCAGATGTGTCAAGAGCAGTGCAGCGACGAGAACCGG 358
QY 121 TTAAAGTGTCACTGTATGTCGAATCTCATCAAGAACAACTGTTGCTGCTTCAGAAAC 180
Db 359 TTCAAGTGCATGTCATGTGCGAGTGCACACGAGGCAATGATCGGTTCGGCATGGCC 418
QY 181 CCTCAGCAGTTTATGGAATATTTTTCAGAGGAATTCGGAATGACTTTCTGGAATCTTG 240
Db 419 CCGACCGCGTGTGTCGAGGGCTTCTCCGAGGAGTTCCTCGAGTTCCTCTCCCTCATC 478
QY 241 AGCGACGCTTTGGCACTAAAGGFTCCACAACATTGTCTACAAATGAATACATCAGC 300
Db 479 CGCCGCGCGCACCGCACTCCGCGTCCGCGCACCGTGTCTACAAAGTACATCGCC 538
QY 301 CACCGAGACCATCCACATGAACGCTACCCAGTGGGAGACACTGACCGACTTTACCAAG 360
Db 539 GACCGGCACACGTCACATCACTCCAGCGGTGGGCCACGCTCACCGAGTTCGTCAAG 598
QY 361 TGGCTGGGCAGAGAGGGCTTGTGTAAG 388
Db 599 TTCCTGGGCGCGAGGGGTACTGTAAAG 467

Query Match 15.3%; Score 168.8; DB 7; Length 1713;
Best Local Similarity 64.7%; Pred. No. 4.8e-26;
Matches 251; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 1 ATGGGCAAGTCGGATTCTTCGAGCCCAAGGCCATCGCCAATAGAAATTAAGTCCAAAGGG 60
Db 239 ATGGGCAAGTCGGATTCTTCGAGCCGGAAGCGATCGCAACCGGATCAAGGCGAAGGGG 298
QY 61 CTCGAGAGCTCGGTGTTACTGCGCAGATGTGCCAAAAGCAATGCCGCGACGAGATGGC 120
Db 299 CTCGAGAGCTCGGTGTTACTGTCAGATGTGTCAAGAGCAGTGCAGCGACGAGAACCGG 358
QY 121 TTAAAGTGTCACTGTATGTCGAATCTCATCAAGAACAACTGTTGCTGCTTCAGAAAC 180
Db 359 TTCAAGTGCATGTCATGTGCGAGTGCACACGAGGCAATGATCGGTTCGGCATGGCC 418
QY 181 CCTCAGCAGTTTATGGAATATTTTTCAGAGGAATTCGGAATGACTTTCTGGAATCTTG 240
Db 419 CCGACCGCGTGTGTCGAGGGCTTCTCCGAGGAGTTCCTCGAGTTCCTCTCCCTCATC 478
QY 241 AGCGACGCTTTGGCACTAAAGGFTCCACAACATTGTCTACAAATGAATACATCAGC 300
Db 479 CGCCGCGCGCACCGCACTCCGCGTCCGCGCACCGTGTCTACAAAGTACATCGCC 538
QY 301 CACCGAGACCATCCACATGAACGCTACCCAGTGGGAGACACTGACCGACTTTACCAAG 360
Db 539 GACCGGCACACGTCACATCACTCCAGCGGTGGGCCACGCTCACCGAGTTCGTCAAG 598
QY 361 TGGCTGGGCAGAGAGGGCTTGTGTAAG 388
Db 599 TTCCTGGGCGCGAGGGGTACTGTAAAG 467
```

RESULT 8

```
US-10-739-930-2955
; Sequence 2955, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(5337)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 2955
; LENGTH: 1713
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: ZEAMA-23APR03-CLUSTER7583_2
US-10-739-930-2955
```

```
Query Match 15.2%; Score 167.6; DB 8; Length 1713;
Best Local Similarity 64.4%; Pred. No. 8.8e-26;
Matches 251; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

QY 1 ATGGCAAGTCGGATTCTTCGAGCCCAAGGCCATCGCCAATAGAAATTAAGTCCAAAGGG 60
Db 175 ATGGGGAAGCAGAGTTCCTTCAGCGCGAAGCGGATCGCGAACCCGGATCAAGGCGAAGGGG 234
QY 61 CTCGAGAGACTTCGCTGTTACTGCCAGATGTGCCAAAAGCAATGCCGCGACGAGATGGC 120
Db 235 CTCGAGAGACTTCGCTGTTACTGCCAGATGTGTCAAGAGCAGTGCCTCGGACGAGAACGGG 294
QY 121 TTAAAGTGTCACTGTATGTCGAATCTCATCAAGAACAACTGTTGCTGCTTCAGAAAC 180
Db 295 TTCAGTGCCTGTCATGTGCGAGTGCACACGAGCGCAGATGCAGGTGTTGGCATGGCG 354
QY 181 CCTCAGCAGTTTATGGAATATTTTTCAGAGGAATTCGGAATGACTTTCTGGAATCTTG 240
Db 355 CCCGACCGCGTGTGTCGAGGGCTTCTCCGAGGAATTCCTCGAGTTCCTCTCCCTCATC 414
QY 241 AGCGCAGCTTTGGCACTAAAGGFTCCACAACAACTGTTCTACAAATGAATACATCAGC 300
Db 415 CGCCGCGCGCACCGCACTCCGCGTCCGCGCACCGTGTCTACAAAGTACATCGCC 474
QY 301 CACCGAGACCATCCACATGAACGCTACCCAGTGGGAGACACTGACCGACTTTACCAAG 360
Db 475 GACCGGCACACGTCACATCACTTACGCGGTGGGCCACGCTCACCGAGTTCGTCAAG 534
QY 361 TGGCTGGGCAGAGAGGGCTTGTGTAAGGT 390
Db 535 CTCCTGGGCGCGAAGGGTACTGTCAAGGTT 564
```

RESULT 9

```
US-10-424-599-25484
; Sequence 25484, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 25484
; LENGTH: 777
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
```

```
; OTHER INFORMATION: Clone ID: PAT_MRT3847_123013C.1
US-10-424-599-25484

Query Match      15.0%; Score 165; DB 7; Length 777;
Best Local Similarity 63.0%; Pred. No. 2.3e-25;
Matches 255; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

QY 1 ATGGGCAAGTCGATTTCTGAGCCCAAGGCCATCGCCAAATAGATTAAATCCAAAGG 60
DB 140 ATGGGCAAAATAGATTCTCACCCTAAGCAATTCGCCATCGAATCAAGCAAAAGGA 199
QY 61 CTCGAGAAGCTTCGCTGGTACTCCAGATGTGCAAAAGCAATGCGCGCAGCAAGATGGC 120
DB 200 TTGCAAGAGCTTCGGTGGTATTGCCAGATGTGTCAGAGCAGTGCAGAGATGAGATGG 259
QY 121 TTTAAGTGCATGTATGTCTGAATCTCATCAAGAACAATGTTGCTGGCTTCAGAAAC 180
DB 260 TTTAAATGCCATTGCCATGATGAGGCAACAGCGTCAATGTCAGATTTTGTGCAAAAC 319
QY 181 CCTCAGCAGTTATGATTTATTTTTCAGAGGAATTCGGAATGACATTTCTGGAATCTG 240
DB 320 CCACACCGGATGTTGAGGCTATTCCGAAGAGTTTGAGAGTACTTTTCTGGAGCATG 379
QY 241 AGCGGAGCTTTGGCACTAAAGGCTCCACAACAACATTTCTTACAATGAATACATCAGC 300
DB 380 AAGCGAGTACCGATTCACCGCTGTGGCAGCCACTGTAGTTTATACGATACATAAAT 439
QY 301 CACCGAGAGCACATCCACATGAACGCTACCCAGTGGGAGACACTGACCGACTTTTACAAG 360
DB 440 GACAGACACCATTCATATGAATCTACTCAGTGGGCTACGCTTACTGAGTTTGTAAAG 499
QY 361 TGGCTGGGAGAGGCTGTGTAAAGGTACCAAGTGCACCTGAAG 405
DB 500 TACTTGGGTCGAATCGCAAAATGTAAGGTTGAGGAAACACCCCAAG 544

RESULT 10
US-11-097-143-41039
; Sequence 41039, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41039
; LENGTH: 1241
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-41039

Query Match      14.8%; Score 163; DB 10; Length 3313;
Best Local Similarity 65.0%; Pred. No. 1.2e-24;
Matches 241; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

; OTHER INFORMATION: Clone ID: PAT_MRT3847_123013C.1
US-10-424-599-25484

Query Match      15.0%; Score 165; DB 7; Length 777;
Best Local Similarity 63.0%; Pred. No. 2.3e-25;
Matches 255; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

QY 1 ATGGGCAAGTCGATTTCTGAGCCCAAGGCCATCGCCAAATAGATTAAATCCAAAGG 60
DB 140 ATGGGCAAAATAGATTCTCACCCTAAGCAATTCGCCATCGAATCAAGCAAAAGGA 199
QY 61 CTCGAGAAGCTTCGCTGGTACTCCAGATGTGCAAAAGCAATGCGCGCAGCAAGATGGC 120
DB 200 TTGCAAGAGCTTCGGTGGTATTGCCAGATGTGTCAGAGCAGTGCAGAGATGAGATGG 259
QY 121 TTTAAGTGCATGTATGTCTGAATCTCATCAAGAACAATGTTGCTGGCTTCAGAAAC 180
DB 260 TTTAAATGCCATTGCCATGATGAGGCAACAGCGTCAATGTCAGATTTTGTGCAAAAC 319
QY 181 CCTCAGCAGTTATGATTTATTTTTCAGAGGAATTCGGAATGACATTTCTGGAATCTG 240
DB 320 CCACACCGGATGTTGAGGCTATTCCGAAGAGTTTGAGAGTACTTTTCTGGAGCATG 379
QY 241 AGCGGAGCTTTGGCACTAAAGGCTCCACAACAACATTTCTTACAATGAATACATCAGC 300
DB 380 AAGCGAGTACCGATTCACCGCTGTGGCAGCCACTGTAGTTTATACGATACATAAAT 439
QY 301 CACCGAGAGCACATCCACATGAACGCTACCCAGTGGGAGACACTGACCGACTTTTACAAG 360
DB 440 GACAGACACCATTCATATGAATCTACTCAGTGGGCTACGCTTACTGAGTTTGTAAAG 499
QY 361 TGGCTGGGAGAGGCTGTGTAAAGGTACCAAGTGCACCTGAAG 405
DB 500 TACTTGGGTCGAATCGCAAAATGTAAGGTTGAGGAAACACCCCAAG 544

RESULT 11
US-11-097-143-41038
; Sequence 41038, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41038
; LENGTH: 3313
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-41038

Query Match      14.8%; Score 163; DB 10; Length 3313;
Best Local Similarity 65.0%; Pred. No. 1.2e-24;
Matches 241; Conservative 0; Mismatches 130; Indels 0; Gaps 0;
```

QY 1 ATGGGCAAGTCGGATTTCTGAGCCCCCAAGGCCATCGCCCAATAGAAATTAAGTCCAAAGGG 60
DB 1069 ATGGGTCGGCGGAGGTAGGTACGCCCAAGTACCTCGCCCAACAAGATGAATCGAAGGTT 1128
QY 61 CTCGAGAGCTTCGGCTGGTACTGCGAGATGTCGCAAAAGCAATGCGGACGAGAAATGCG 120
DB 1129 CTGCAAGAGCTCGGCTGGTACTGCGCAGATGTGCGAAGCAATGCGCGCATGAACCGGC 1188
QY 121 TTTAAGTGTCACTGTATGTCTGAATCTCATCAAGACAACTGTGTGCTGCTTCAGAAAC 180
DB 1189 TTTCAAGTGCACAGATGAGCGAGTCCACAGCGCCAGTTGCTCTCTTTGGGACAAAT 1248
QY 181 CTTCAAGAGTTATGGAATTTATTTTCAGAGGAATTCGCAAAATGACTTTCTGGAATCTTCTG 240
DB 1249 CTTGCAAAATCTCTGCACAGCTTCAGCAAAAGAGTTCTCCGAGGGCTACATGGAATGTTG 1308
QY 241 AGCGGAGCTTTGGCACTAAAGGGTCCACAACAACATTTGTCTACAAATGAATACATCAGC 300
DB 1309 CGCCGGCGGTTTCGGCACAGAGGGAACCGAGCGCCAAAGATCTTACGAGGTACATTTGCC 1368
QY 301 CACCGAGAGCACATCCACATGAACGCTACCCAGTGGGAGACATGACCGCACTTTTACCAG 360
DB 1369 CACAAGGAGCACATCCACATGAAGCCACCCGATGGCTCACCTGTCCGACTACGTGAAG 1428
QY 361 TGGCTGGGCAG 371
DB 1429 TGGCTGGGCG 1439

RESULT 12

US-11-097-143-41017
; Sequence 41017, Application US/11097143
; Publication No. US2005020858A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: DROSOPHILA GENES.
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41017
; LENGTH: 4394
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-41017

Query Match 14.8%; Score 163; DB 10; Length 4394;
Best Local Similarity 65.0%; Pred. No. 1.3e-24;
Matches 241; Conservative 0; Mismatches 130; Indels 0; Gaps 0;
QY 1 ATGGGCAAGTCGGATTTCTGAGCCCCCAAGGCCATCGCCCAATAGAAATTAAGTCCAAAGGG 60

DB 3758 ATGGTCCGCGGAGGTAGGTACGCCCAAGTACCTCGCCCAACAAGATGAATCGAAGGTT 3817
QY 61 CTCGAGAGCTTCGGCTGGTACTGCGAGATGTCGCAAAAGCAATGCGGCGACGAGAAATGCG 120
DB 3818 CTGCAAGAGCTCGGCTGGTACTGCGCAGATGTGCGAAGCAATGCGCGCATGAACCGGC 3877
QY 121 TTTAAGTGTCACTGTATGTCTGAATCTCATCAAGACAACTGTGTGCTGCTTCAGAAAC 180
DB 3878 TTTCAAGTGCACAGATGAGCGAGTCCACAGCGCCAGTTGCTCTCTTTGGGACAAAT 3937
QY 181 CTTCAAGAGTTATGGAATTTATTTTCAGAGGAATTCGAAATGACTTTCTGGAATCTTCTG 240
DB 3938 CTTGCAAAATCTCTGCACAGCTTCAGCAAAAGAGTTCTCCGAGCGCTACATGGAATGCTG 3997
QY 241 AGCGGAGCTTTGGCACTAAAGGGTCCACAACAACATTTGTCTACAAATGAATACATCAGC 300
DB 3998 CGCCGGCGGTTTCGGCACAGAGGGAACCGAGCGCCAAAGATCTTACGAGGTACATTTGCC 4057
QY 301 CACCGAGAGCACATCCACATGAACGCTACCCAGTGGGAGACATGACCGCACTTTTACCAG 360
DB 4058 CACAAGGAGCACATCCACATGAAGCCACCCGATGGCTCACCTGTCCGACTACGTGAAG 4117
QY 361 TGGCTGGGCAG 371
DB 4118 TGGCTGGGCG 4128

RESULT 13

US-11-097-143-24532
; Sequence 24532, Application US/11097143
; Publication No. US2005020858A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24532
; LENGTH: 4582
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-24532

Query Match 14.8%; Score 163; DB 10; Length 4582;
Best Local Similarity 65.0%; Pred. No. 1.3e-24;
Matches 241; Conservative 0; Mismatches 130; Indels 0; Gaps 0;
QY 1 ATGGGCAAGTCGGATTTCTGAGCCCCCAAGGCCATCGCCCAATAGAAATTAAGTCCAAAGGG 60
DB 3946 ATGGGTCGCGCGAGGTAGGTACGCCCAAGTACCTCGCCCAACAAGATGAATCGAAGGTT 4005
QY 61 CTCGAGAGCTTCGGCTGGTACTGCGCCAAAGCAATGCGCGACGAGAAATGCG 120

Db 4006 CTCAGAGCTGCGCTGGTACTCCAGATGTGCGAGAGCAGTGCAGGATGAACGGC 4065
Qy 121 TTAAAGTGTCACTGTATGTCTGAATCTCATCAAGACAAGTGTGCTGGCTTCAGAAAAC 180
Db 4066 TTCAAGTGCACACGATGAGCGAGTCCACAGCGCAGTGTCTCTCTTTGCGGACAAT 4125
Qy 181 CCTCAGCAGTTATGATATTTTTCAGAGGAATTCGAAATGACATTTCTGGAACCTTCTG 240
Db 4126 CCTGGCAAAATTCCTGCACAGCTTCAGCAAGAGTTCCTCGACGGCTACATGGAGTTGCTG 4185
Qy 241 AGCGAGCTTTGGCACTAAAGGTTCCACAACAATTTCTTACAAATGAATACATCAGC 300
Db 4186 CGCGGGGGTTGGCAGCAGCGAACCAGCGCCACAGATCTACAGGAGTACATGGCC 4245
Qy 301 CACGAGAGCAGATCCACATGAACGCTACCCAGTGGGAGACACTGACCGACTTTACCAAG 360
Db 4246 CACAAGGAGCAGATCCACATGAACGCGCACCGATGGCTCACCTGTCCGACTACGTGAAG 4305
Qy 361 TGGCTGGGCGAG 371
Db 4306 TGGCTGGGCGG 4316

RESULT 14

US-10-767-701-25587
; Sequence 25587, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 53128
; SEQ ID NO 25587
; LENGTH: 584
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 30963818
US-10-767-701-25587

Query Match 13.6%; Score 150.4; DB 7; Length 584;
Best Local Similarity 63.6%; Pred. No. 2.8e-22;
Matches 229; Conservative 0; Mismatches 131; Indels 0; Gaps 0;
Qy 1 ATGGGCAAGTCGAGTTTCTGAGCCCCCAAGGCCATCGCCCAATAGAAATTAAGTCCAAAGGG 60
Db 225 ATGGGCAAGTCGAGTTTCTGAGCGGAGGCGATCGGCAACCGGATCAAGGCAAAAGGG 284
Qy 61 CTCGAGAGCTTCGCTGTACTGCGAGATGTGCCAAGCAATGCCGCGAGAGATGCG 120
Db 285 CTGCAAAAGCTGCGGTGTACTGCGAGATGTGCCAAGCAATGCCGCGAGAGATGCG 344
Qy 121 TTAAAGTGTCACTGTATGTCTGAATCTCATCAAGACAATCTGTTCTGGCTTCAGAAAAC 180
Db 345 TTCAAGTGTCACTGTATGTCTGAATCTCATCAAGACAATCTGTTCTGGCTTCAGAAAAC 404
Qy 181 CCTCAGCAGTTTATGGATTAATTTTTCAGAGGAATTCGAAATGACTTTCTGGAATTTCTG 240
Db 405 CCCGACCGGCTGCTGAGGGCTTCTCCGAGGATTCCTCGAGTCTCTCTCTCCCTCATC 464
Qy 241 AGCGACGCTTTGGCACTAAAGGGTCCACAACAATTTGCTACAAATGAATACATCAGC 300
Db 465 CGCGCGCGGCAACCGGCACTCCCGCGTCCCGCGCACCGTCTGCTCAACAGAGTACATCGG 524
Qy 301 CACCGAGAGCAGATCCACATGAACGCTACCGAGTGGGAGACACTGACCGACTTTTACCAAG 360
Db 525 GACCGTACCAAGTCCACATGAATCCAGCGCTGGGCGCACCGCTCACCAGTTCGTCAG 584

RESULT 15

US-09-991-936-1424
; Sequence 1424, Application US/09991936
; Publication No. US20030073827A1
; GENERAL INFORMATION:
; APPLICANT: Brandt, Kevin S.
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Stinchcomb, Dan T.
; APPLICANT: Wisniewski, Nancy
; TITLE OF INVENTION: FLEA HEAD, NERVE CORD, HINDGUT AND MALPIGHIAN TUBULE
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES, PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-6-C1
; CURRENT APPLICATION NUMBER: US/09/991,936
; CURRENT FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US/09/543,668
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: 60/128,704
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 1959
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1424
; LENGTH: 549
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
US-09-991-936-1424

Query Match 11.9%; Score 130.8; DB 3; Length 549;
Best Local Similarity 70.7%; Pred. No. 4.6e-18;
Matches 188; Conservative 0; Mismatches 77; Indels 1; Gaps 1;
Qy 1 ATGGGCAAGTCGAGTTTCTGAGCCCCCAAGGCCATCGCCCAATAGAAATTAAGTCCAAAGGG 60
Db 176 ATGGGTAAGCAGAGTAGTAGTACCCCAAGTACATAGCCCAACAAATGAAGGCCAAAGGC 235
Qy 61 CTCGAGAGCTTCGCTGTACTGCCAGATGTGCCAAGCAATGCCCGGAGAGATGCG 120
Db 236 CTCGAGAGCTTCGCTGTACTGCCAATGTGTGAGAAACAATGCAGAGATGAAGACGGT 295
Qy 121 TTAAAGTGTCACTGTATGTCTGAATCTCATCAAGACAATCTGTTCTGGCTTCAGAAAAC 180
Db 296 TTCAATGCCACAAATGTCCGAATCCCAATCCCAAGACAATTTAATCTTTGCTGATAC 355
Qy 181 CCTCAGCAGTTTATGGATTAATTTTTCAGAGGAATTCGAAATGACTTTCTGGAATTTCTG 240
Db 356 TCACACATGTACCTAGATCAATTTTCAAGAGATTTCTGCGGCTTCTTAGAATTTCTG 414
Qy 241 AGCGACGCTTTGGCACTAAAGGGT 266
Db 415 AGAAGACAATTTGGAACCTAAACGAGT 440

Search completed: November 27, 2005, 02:22:26
Job time : 1272.42 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

QM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:41:08 ; Search time 388.941 Seconds
(without alignments)
421.056 Million cell updates/sec

Title: US-09-555-529-2
Perfect score: 1102
Sequence: 1 atggcgaatcggtttctt.....aaaaaaaaaaaaaaaaaaaa 1102

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3205263 seqs, 74304013 residues

Total number of hits satisfying chosen parameters: 6410526

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA New.*
1: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
2: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq2.*
9: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq3.*
10: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48.2	4.4	2120	1	US-10-689-742-163
2	48.2	4.4	2933	1	US-10-131-826A-345
3	47.4	4.3	3501	1	US-10-131-826A-37
4	46.8	4.2	1257	1	US-10-689-742-115
5	46	4.2	1624	1	US-10-131-826A-181
6	45.8	4.2	1108	1	US-10-689-742-209
7	45.6	4.1	1425	1	US-10-689-742-131
8	44.6	4.0	894	1	US-10-689-742-5
9	43.6	3.9	2487	1	US-10-689-742-165
10	43.4	3.9	791	1	US-10-689-742-145
11	43.2	3.9	3819	1	US-10-131-826A-405
12	43.2	3.9	1325	1	US-10-689-742-105
13	42.8	3.9	2033	1	US-10-131-826A-307
14	42.8	3.9	34000	7	US-11-102-978-3
15	42.6	3.8	1523	1	US-10-131-826A-429
16	42	3.8	1658	7	US-11-034-569-15
17	41.8	3.8	2476	1	US-10-131-826A-489
18	41.8	3.8	1904	1	US-10-131-826A-99
19	41.6	3.7	110	1	US-10-689-742-195
20	41.2	3.7	1240	1	US-10-131-826A-91
21	41.2	3.7	4053	1	US-10-131-826A-351
22	41.2	3.7	2066	1	US-10-523-588-2
23	41	3.7			

ALIGNMENTS

RESULT 1
US-10-689-742-163
; Sequence 163, Application US/10689742
; Publication No. US20050250180A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M
; APPLICANT: LaVallie, Edward R
; APPLICANT: Racie, Lisa A
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 00766.000091.10
; CURRENT APPLICATION NUMBER: US/10/689,742
; CURRENT FILING DATE: 2003-10-22
; PRIOR APPLICATION NUMBER: 09/746,783
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 163
; LENGTH: 2120
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-689-742-163
Query Match 4.4%; Score 48.2; DB 1; Length 2120;
Best Local Similarity 51.6%; Pred. No. 0.059;
Matches 110; Conservative 0; Mismatches 103; Indels 0; Gaps 0;
Qy 890 TTTGAAATTTGATAACAACATTTGAACTGTGAAGCATCAATTTGGTGTAGCCCAAGG 949
Db 1866 TTTTAAATACCATATCATGCTATGTAATAATAAGACTACTGGCTTTGTGAGGGAATGT 1925
Qy 950 CACTGTGTAATCTACTGTGTAGGGATTTGTTTGTATTAATAAAAAAATCATCT 1009
Db 1926 TTGTGCAAAATTTTTCCTCTAATATGTTAAATTTGATTAATAATCTTCAGA 1985
Qy 1010 ATTAAATATCTAGTGAATAGTTGGTAAATTTTAAATAAATCTATGTTTTTAAAGTG 1069
Db 1986 ATTAATATCCCTTTTGTCTACTTTTGAACACATATAATCATTTGTAATCTGTCCTTA 2045
Qy 1070 TAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1102
Db 2046 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2078

```
RESULT 2
US-10-131-826A-345
; Sequence 345, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330RIC128
; CURRENT APPLICATION NUMBER: US/10/131,826A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 345
; LENGTH: 2933
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-131-826A-345

Query Match 4.4%; Score 48.2; DB 1; Length 2933;
Best Local Similarity 58.9%; Pred. No. 0.066;
Matches 83; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 962 CTACTGTGTTAGGGATTTGTTTGTATTAATAAATCATCTATTTAAATACTA 1021
Db 2791 CAAATGTTATTCAGTTTCTTACTGTTTGTAAATGAGAAATTCCTTTTAAATATTT 2850
QY 1022 GTGATAGTGGCTAATTTATTAATAAATCATCTATTTTAACTGTAAATAAATAA 1081
Db 2851 TTCCAAATAAATTTTATGAAATGACAAAAAATAAATAAATAAATAAATAAATAA 2910
QY 1082 AAAAAAAAAAAAAAAAAAAAAA 1102
Db 2911 AAAAAAAAAAAAAAAAAAAAAA 2931

RESULT 3
US-10-131-826A-37
; Sequence 37, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330RIC128
; CURRENT APPLICATION NUMBER: US/10/131,826A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 37
; LENGTH: 3501
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 2762, 2778
; OTHER INFORMATION: unknown base
US-10-131-826A-37

Query Match 4.3%; Score 47.4; DB 1; Length 3501;
Best Local Similarity 62.0%; Pred. No. 0.11;
Matches 75; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 982 TTTTCTATTAAAAAATAATCATCTATTTAAATACCTAGTGAATAGTTGGTAAATTT 1041
Db 3379 TTTTCTTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3438
QY 1042 ATAATAAATCTATCTGTTTTTTTTTAACTGTAAATAAATAAATAAATAAATAA 1101
Db 3439 AAAAAAAAAAAAAAAAAAAAAA 3498
QY 1102 A 1102
```

```
Db      3499 A 3499

RESULT 4
US-10-689-742-115
; Sequence 115, Application US/10689742
; Publication No. US20050250180A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M
; APPLICANT: Lavallie, Edward R
; APPLICANT: Racie, Lisa A
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 00766.000091.10
; CURRENT APPLICATION NUMBER: US/10/689,742
; CURRENT FILING DATE: 2003-10-22
; PRIOR APPLICATION NUMBER: 09/746,783
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 115
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-689-742-115

Query Match      4.2%; Score 46.8; DB 1; Length 1257;
Best Local Similarity 73.2%; Pred. No. 0.097;
Matches 60; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 1021 AGTGAATAGTTGGTAAATTTATAATAAAATCTATGTTTTTTTAAAGTGTGTAATAAAAAA 1080
Db 1172 AGTAGTAATTTAAAGAACTCAATAAAACCTCTATTTTTTATTTTAAATAATATAAAAAA 1231
QY 1081 AAAAAAAAAAAAAAAAAAAAAA 1102
Db 1232 AAAAAAAAAAAAAAAAAAAAAA 1253

RESULT 5
US-10-131-826A-181
; Sequence 181, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C128
; CURRENT APPLICATION NUMBER: US/10/131,826A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18

Db      3499 A 3499

PRIORITY APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 181
; LENGTH: 1624
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 1560-1561
; OTHER INFORMATION: unknown base
US-10-131-826A-181

Query Match      4.2%; Score 46; DB 1; Length 1624;
Best Local Similarity 58.1%; Pred. No. 0.16;
Matches 79; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 967 GTGTTAGGGGATTTGTTTGTATTAATAAAAAAATCATCTATTTAATATCTAGTGAA 1026
Db 1469 GTTGATTCCTATTTTGTAGCTTGCAAAAAGTGACTTATATTCCAAGAAATTAATAATGTT 1528
QY 1027 TAGTGGGTAAATTTATATAAATCTATGTTTTTTTAAAGTGTAAATAAAAAA 1086
Db 1529 GAAATCCAAATCCTAGAAATAAATAGTTTNNTTCCAAAAAATAAAAAA 1588
QY 1087 AAAAAAAAAAAAAAAAAA 1102
Db 1589 AAAAAAAAAAAAAAAAAA 1604

RESULT 6
US-10-689-742-209
; Sequence 209, Application US/10689742
; Publication No. US20050250180A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M
; APPLICANT: Lavallie, Edward R
; APPLICANT: Racie, Lisa A
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 00766.000091.10
; CURRENT APPLICATION NUMBER: US/10/689,742
; CURRENT FILING DATE: 2003-10-22
; PRIOR APPLICATION NUMBER: 09/746,783
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 209
; LENGTH: 1108
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-689-742-209
```

```
Query Match          4.2%; Score 45.8; DB 1; Length 1108;
Best Local Similarity 61.2%; Pred. No. 0.15;
Matches 74; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 982 TTTTGTATTAAAAAATCACTCTATTAAATACCTAGTGAATGTTGGTAAATTT 1041
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 940 TTTATTAAAAAAGAAAAAAGAAAAAAGAGGGAGGAAAAAAGAACCTACTT 999
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1042 ATAATAAAATCTATGTTTTTTTAAAGTGTAATAAAAAAAAAAAAAAAAAA 1101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1000 AAGATAGAACATGGTTTGGCTATGCTTGAATAAAAAAAAAAAAAAAAAA 1059
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1102 A 1102
DB 1060 A 1060

RESULT 7
US-10-689-742-131
; Sequence 131, Application US/10689742
; Publication No. US20050250180A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M
; APPLICANT: LaVallie, Edward R
; APPLICANT: Racie, Lisa A
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 00766.000091.10
; CURRENT APPLICATION NUMBER: US/10/689,742
; PRIOR FILING DATE: 2003-10-22
; PRIOR APPLICATION NUMBER: 09/746,783
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 131
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-689-742-131

Query Match          4.1%; Score 45.6; DB 1; Length 1425;
Best Local Similarity 71.4%; Pred. No. 0.19;
Matches 60; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 1019 CTAGTGAATAGTGGTAAATTTAATAAATCTATGTTTTTTTAAAGTGA 1078
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1339 CTGGAACATGGATGGGGATTGATAGAAAAATAAACCCCTGCTTTTGAT 1398
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1079 AAAAAAAAAAAAAAAAAAAAAA 1102
DB 1399 AAAAAAAAAAAAAAAAAAAAAA 1422

RESULT 8
US-10-689-742-41
; Sequence 41, Application US/10689742
; Publication No. US20050250180A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M
; APPLICANT: LaVallie, Edward R
; APPLICANT: Racie, Lisa A
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 00766.000091.10

; CURRENT APPLICATION NUMBER: US/10/689,742
; PRIOR FILING DATE: 2003-10-22
; PRIOR APPLICATION NUMBER: 09/746,783
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 41
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-689-742-41

Query Match          4.0%; Score 44.6; DB 1; Length 588;
Best Local Similarity 53.1%; Pred. No. 0.22;
Matches 95; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 924 AAGCATCAAAATTTGGTGTAGCCCAAGGCACCTGTAACCTCTACTGTGTAG 983
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 356 AAGAAACCAAGAGGCTAATCACAGGGTTCGAAATAAATAAATCTGATTAG 415
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 984 TTGTATTAAAAAATAAATCATCTATTATAACTAGTGAATAGTTGGGTAAT 1043
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 416 TAGGAATTAATAAATGAAACCTCATGTAAATTTAAGATAGATTCAAACAT 475
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1044 AATAAAATCTATGTTTTTTTAAAGTGTAATAAAAAAAAAAAAAAAAAA 1102
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 476 AAGGATAGCAAGAATTTACACAAAAAATAAATAAATAAATAAATAAATAA 534
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
US-10-689-742-5
; Sequence 5, Application US/10689742
; Publication No. US20050250180A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M
; APPLICANT: LaVallie, Edward R
; APPLICANT: Racie, Lisa A
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 00766.000091.10
; CURRENT APPLICATION NUMBER: US/10/689,742
; PRIOR FILING DATE: 2003-10-22
; PRIOR APPLICATION NUMBER: 09/746,783
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-689-742-5

Query Match          4.0%; Score 43.6; DB 1; Length 894;
Best Local Similarity 78.8%; Pred. No. 0.42;
Matches 52; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1037 AATTATATAATAAATCTATGTTTTTTTAAAGTGTAAAAAATAAATAA 1096
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 826 AAAATGCAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 885
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1097 AAAAAA 1102
DB 886 AAAAAA 891

RESULT 10
US-10-689-742-165
; Sequence 165, Application US/10689742
; Publication No. US20050250180A1
```

DB 752 AAAAAAAAAAAAAAAAAAAAAA 779

RESULT 12
US-10-131-826A-405
; Sequence 405, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME
; FILE REFERENCE: F333ORAC128
; CURRENT APPLICATION NUMBER: US/10/131,826A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 405
; LENGTH: 3819
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-131-826A-405

Query Match 3.9%; Score 43.2; DB 1; Length 3819;
Best Local Similarity 61.6%; Pred. No. 0.88;
Matches 69, Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy	991	AAAAAAAAAAATCATCTATTATAACTAGTGTGGAATTTTATAATAAAA
Db	3704	AGAAAAATAAAAAATCAGCTGTCGTAATCGCTAGCAAAAAAAAAAAAAAAAAAAAAA
Qy	1051	TCTATGTTTTTTTAGTGTAIAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
Db	3764	AA

RESULT 13
US-10-689-742-105

```
; Sequence 105, Application US/10689742
; Publication No. US20050250180A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M
; APPLICANT: Lavallie, Edward R
; APPLICANT: Racie, Lisa A
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 00766.00091.10
; CURRENT APPLICATION NUMBER: US/10/689,742
; CURRENT FILING DATE: 2003-10-22
; PRIOR APPLICATION NUMBER: 09/746,783
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 105
; LENGTH: 1325
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-689-742-105

Query Match      3.9%; Score 42.8; DB 1; Length 1325;
Best Local Similarity 61.6%; Pred. No. 0.73;
Matches 85; Conservative 0; Mismatches 52; Indels 1; Gaps 1;

QY 966 TCGTGTAGGGGATTTGTTTGTATTAAAAAATCACTATTTAACTACTA-GTG 1024
Db 1151 TTTTGTAGTCAATGAACCTTTTGTCAAAAACAGTAAACATCTGTGTTTAACTATGGT 1210
QY 1025 AATAGTTGGGTAAATTTATAATAAAATCTATGTTTTTTTAAAGTGTAAAAA 1084
Db 1211 AAACAACATGTTAATGAACATGCTATCCATGACTTAATGGACAGTTCAAAAA 1270
QY 1085 AAAAAA 1102
Db 1271 AAAAAA 1288
```

```
RESULT 14
US-10-131-826A-307
; Sequence 307, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C128
; CURRENT APPLICATION NUMBER: US/10/131,826A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
```

```
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 307
; LENGTH: 2033
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-131-826A-307

Query Match      3.9%; Score 42.8; DB 1; Length 2033;
Best Local Similarity 58.7%; Pred. No. 0.85;
Matches 74; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 970 TTAGGGGATTTGTTTGTATTAAAAAATCACTATTTAACTACTAGTGAATAG 1029
Db 1908 TGAAGCAATATATATATTTGTAAACAAAACAGCTCTTACCTATAAACATTTTATACT 1967
QY 1030 TTGGTAAATTTATAATAAAATCTATGTTTTTTTAACTGTAAAAA 1089
Db 1968 GTTGTATGTATAAATAAAGTGTCTTGTAGTTTTTTTGGAAAAA 2027
QY 1090 AAAAAA 1095
Db 2028 AAAAAA 2033
```

```
RESULT 15
US-11-102-978-3
; Sequence 3, Application US/11102978
; Publication No. US20050250142A1
; GENERAL INFORMATION:
; APPLICANT: University of Utah Technology Transfer Office
; APPLICANT: University of Utah Research Foundation
; TITLE OF INVENTION: Diagnosis and Treatment of Herpes Simplex Virus Disease
; FILE REFERENCE: 0274-5537.1US
; CURRENT APPLICATION NUMBER: US/11/102,978
; CURRENT FILING DATE: 2005-04-11
; PRIOR APPLICATION NUMBER: PCT/US2003/033152
; PRIOR FILING DATE: 2003-10-18
; PRIOR APPLICATION NUMBER: 60/419,576
; PRIOR FILING DATE: 2002-10-18
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 340000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: (56948)..(57115)
; OTHER INFORMATION: C21orf34 exon
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (80006)..(81089)
; OTHER INFORMATION: Gene VDACC2P; voltage-dependent anion channel isoform 2 pseudogene
; FEATURE:
; NAME/KEY: exon
; LOCATION: (167308)..(167438)
```


; OTHER INFORMATION: C21orf34 exon
; FEATURE:
; NAME/KEY: exon
; LOCATION: (216732)..(216833)
; OTHER INFORMATION: C21orf34 exon
US-11-102-978-3

Query Match 3.9%; Score 42.6; DB 7; Length 340000;
Best Local Similarity 47.9%; Pred. No. 6.3; Indels 0; Gaps 0;
Matches 123; Conservative 0; Mismatches 134;
QY 846 AGTTGAAGGATTCATATGAGACATATCTAACTTGGCTTGAGTTTGAAAAATTGATAA 905
Db 301737 AGTCAATACATATAGTTTGACTATATATAGTTTATTCATATATAGGATATATGAATAT 301796
QY 906 CAACACATTGAACTGTGAAGCATCAAATGGTGTAGCCAGGCACTGTGTAACCTCTAC 965
Db 301797 ATAAATATAGATATATTTTATTCATATATATCCTATATATAGAAATATATAACTATAT 301856
QY 966 TGTGTTAGGGGATTTGTTTGTATTAATAAAAAAATCATCTATTAAATCTAGTGA 1025
Db 301857 GTAAGCCAAATATATATAAAAGTAACAAATATATTTTACTATATATTTGAATATATAAAT 301916
QY 1026 ATAGTTGGGTAATTTATATAAAATCTATGTTTGTAAAGTGTAAAAAATAAAAAA 1085
Db 301917 GTATTTGAATATATATGTTTGAATATATAAATTTATGCAAAATAAAACATATAATAA 301976
QY 1086 AAAAAAAAAAAAAA 1102
Db 301977 ATATATATAGTCAATA 301993

Search completed: November 27, 2005, 02:44:29
Job time : 390.941 secs

This page Blank (used)

181 AAGTGTCAATGATGTCGCAATCTCATCAGAGACAACTATTGCTGGCTTCAGAAATCCT 240
181 AAGTGTCAATGATGTCGCAATCTCATCAGAGACAACTATTGCTGGCTTCAGAAATCCT 240
241 CAGCAGTTTATGGATTAATTTTTCAGAGGAATTCGGAATGACTTTCTAGAACTTCTCAGG 300
241 CAGCAGTTTATGGATTAATTTTTCAGAGGAATTCGGAATGACTTTCTAGAACTTCTCAGG 300
301 AGACGCTTTGGCACTAAAGGGTCCAAACAACATTTGTCTACAAAGAAATACATCAGCCAC 360
301 AGACGCTTTGGCACTAAAGGGTCCAAACAACATTTGTCTACAAAGAAATACATCAGCCAC 360
361 CGAGAGCACATCCATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
361 CGAGAGCACATCCATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
421 CTGGGCGAGAGAGGCTTTGTGCAAAAGTGCATGAGACGATAGGAAGTTTCAGCATCAGTG 480
421 CTGGGCGAGAGAGGCTTTGTGCAAAAGTGCATGAGACGATAGGAAGTTTCAGCATCAGTG 480
481 AAACGAAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAGAAAGAAAGAAATCT 540
481 AAACGAAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAGAAAGAAAGAAATCT 540
541 GCACCTGGATGAAATCATGAGATTTGAAGAGGAAAGAAAGAAAGAAAGAAATCT 600
541 GCACCTGGATGAAATCATGAGATTTGAAGAGGAAAGAAAGAAAGAAAGAAATCT 600
601 TGCTACAGCTGAAATTAATTTGAAATTTAATCAAGAAAGTGGGAGAGAAATATCAT 660
601 TGCTACAGCTGAAATTAATTTGAAATTTAATCAAGAAAGTGGGAGAGAAATATCAT 660
661 AAGAAAAGGCTATTGTTAAGGAAGTAATGACAAATATACAGCTGTTGTGAAGATGATT 720
661 AAGAAAAGGCTATTGTTAAGGAAGTAATGACAAATATACAGCTGTTGTGAAGATGATT 720
721 GATTCGAGACAGCTGAACTTGACACAGCTCATTTAGACACAGTAATTCAGGACCA 780
721 GATTCGAGACAGCTGAACTTGACACAGCTCATTTAGACACAGTAATTCAGGACCA 780
781 GGAAGAAAGAAATCTAGTTTAAATGAGGCTACAGAGGAATGAAGTACCCTAGAAATCC 840
781 GGAAGAAAGAAATCTAGTTTAAATGAGGCTACAGAGGAATGAAGTACCCTAGAAATCC 840
841 ATCAATGAGAGAGCTTTTTCAGCTACTATCGTCAATGAAACTGGCCCTTTAAAGAGAGC 900
841 ATCAATGAGAGAGCTTTTTCAGCTACTATCGTCAATGAAACTGGCCCTTTAAAGAGAGC 900
901 AGAGTTGAAGGAATTCATATGAGACATTTCTTAACTTGCCTGAGTTTGAAGAAATTTGTT 960
901 AGAGTTGAAGGAATTCATATGAGACATTTCTTAACTTGCCTGAGTTTGAAGAAATTTGTT 960
961 AACAATACCTTTAAATCTTTAAGCATCAATTTGGTTCGC 1002
961 AACAATACCTTTAAATCTTTAAGCATCAATTTGGTTCGC 1002

RESULT 2
AX003309
LOCUS AX003309 1102 bp DNA linear PAT 24-AUG-2000
DEFINITION Sequence 2 from Patent WO9929845.
ACCESSION AX003309
VERSION AX003309.1 GI:9927126
KEYWORDS
SOURCE Mus sp.
ORGANISM Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1
Angulo-Mora, J.F. and Mauffrey, P.
AUTHORS Sequences coding for kin17 protein and their applications
TITLE Patent: WO 9929845-A 2 17-JUN-1999,
JOURNAL

ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE (FR)
FEATURES
source Location/Qualifiers
1..1102
/organism="Mus sp."
/mol_type="unassigned DNA"
/db_xref="taxon:10095"
ORIGIN
Query Match 70.3%; Score 704.6; DB 6; Length 1102;
Best Local Similarity 86.1%; Pred. No. 1.4e-156;
Matches 819; Conservative 0; Mismatches 119; Indels 13; Gaps 3;
QY 58 ATGGGGAAGTCGGATTTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCCAAGGG 117
DB 1 ATGGGCAAGTCGGATTTTCTGAGCCCCCAAGGCCAATAGAAATTAAGTCCCAAGGG 60
QY 118 CTGCAGAAAGCTACGCTGGTATTGGCAGATGTGCCAAGAGCAGTGCCTGGACAGAAATGCG 177
DB 61 CTCGAGAAGCTTCGCTGGTACTGCGAGATGTGCCAAGAGCAATGCCGCGACGAGAAATGCG 120
QY 178 TTTAAGTGTCAATGATGTCGGAATCTCATCAGAGACAACTATTGCTGGCTTCAGAAAT 237
DB 121 TTTAAGTGTCAATGATGTCGGAATCTCATCAAGAGCAACTGTTGCTGGCTTCAGAAAT 180
QY 238 CCTCAGAGATTTATGGATTTATTTTCAGAGGAATTCGGAATGACTTTCTAGAACTTCTC 297
DB 181 CCTCAGAGATTTATGGATTTATTTTCAGAGGAATTCGGAATGACTTTCTGGAATCTTCTG 240
QY 298 AGGAGACGCTTTGGCAGCTAAAGGGTCCCAACAACATTTGTCTACAAAGAAATACATCAGC 357
DB 241 AGGAGACGCTTTGGCAGCTAAAGGGTCCCAACAACATTTGTCTACAAAGAAATACATCAGC 300
QY 358 CACCGAGAGCAGTCCATCAATGAATGCCACTCAGTGGGAAACTCTGAGTGAATTTTACTAAG 417
DB 301 CACCGAGAGCAGTCCATCAATGAATGCCACTCAGTGGGAGACACTGACCGAGCTTTTACCAAG 360
QY 418 TGGCTGGGCGAGAGAGGCTTTGTCMAA-----AGTCAGCTGAAGACATAGGAAGTTCA 471
DB 361 TGGCTGGGCGAGAGAGGCTTTGTCMAAGGTACCAGTGCATGGAAGTCTGCTGGGAGCGCA 420
QY 472 GCATCAGTGAACGAGAAAGAAATCTTCCAGAGCTCAACTCAGTCTTAAAGAAAGAAAGAA 531
DB 421 GCATCCGGGAAACGGAAGAGTCTTTCAGAGAGCTCCGCGCAGCCT-----CGGAAGAG 474
QY 532 AAGAAATCTGCACCTGGATGAATCATGGAGATTGAAGAGGAAAGAAAGAAAGAAAGAAAG 591
DB 475 AAGAAATCTGCACCTGGATGAATCATGGAGCTCGAAGAGGAAAGAAAGAAAGAAAGAAAG 534
QY 592 ACAGACTACTGGCTACAGCCTGAAATTAATTTGTGAAATTAATTAACCAAGAAAGTGGAGAG 651
DB 535 ACAGACGCTCGTTTACAGCCGGGGATCGTTGTGAAATTAATTAACCAAGAAAGTGGGGAG 594
QY 652 AAATATCATAGAAAGAAAGGCTATTGTTAAGGAAGTAAATTTGAAATTAATTAACCAAGAA 711
DB 595 AAATATCATAGAAAGAAAGGCTCGTTAAGGAAGTGAATTTGACAGGTACACAGCTGTGTA 654
QY 712 AAGATGATTGATTTCTGGAGACAAAGCTGAAACTTTGACAGACTCAATTTTGAAGACAGTAAT 771
DB 655 AAGATGATGATGCTCTGGAGACAGGCTGAAACTTGGACACAGACTCAATTTAGAGACAGTCA 714
QY 772 CCAGCACGAGGAAAGAAATTTCTAGTTTTTAATGGAGGCTACAGAGGAATTAAGAGGTACC 831
DB 715 CCGGCCCCGGGAAAGAGGTTCTAGTTTTTAATGGAGGCTACAGAGGAATTAAGAGGCACT 774
QY 832 CTAGAAATCCATCAATGAGAGAACTTTTTTTCAGCTACTATCGTCAATTTGAAAGTGGCCCTTTA 891
DB 775 CTCGAATCAATCAATGAGAGAGCTTTTTTCAGCCAGATAGTCAATTTGAAAGTGGACTTTG 834
QY 892 AAAGGACGAGAGTTGAAGGAATTCATATGAAGACATTTCTTAACTTGAAGTGGCTTTGA 951
DB 835 AAAGGACGAGAGTTGAAGGATTTCAATATGAAGACATATCTAAACTTGTCTTGAAGTTGA 894

```

Qy 952 AAATTTGTTAAACATACCTTTAAATCTTAAAGCATCAAAATTTGTTGTCG 1002
Db 895 AAATTTGATGAACACATCTGAATCT-GTGAAGCATCAAAATTTGTTGTTAGC 944

RESULT 3
AX003308
LOCUS AX003308 1528 bp DNA linear PAT 24-AUG-2000
DEFINITION Sequence 1 from Patent WO929845.
ACCESSION AX003308
VERSION AX003308.1 GI:9927125
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1
AUTHORS Angulo-Mora, J. P. and Mauffrey, P.
TITLE Sequences coding for kin17 protein and their applications
JOURNAL Patent: WO 929845-A 1 17-JUN-1999;
(FR) ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE
FEATURES
source Location/Qualifiers
1..1528
/organism="Homo sapiens"
/db_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 65.2%; Score 653.4; DB 6; Length 1528;
Best Local Similarity 76.6%; Pred. No. 2e-144; Indels 295; Gaps 2;
Matches 969; Conservative 0; Mismatches 1;

Qy 32 AGAAAGTGATCGCTGCGCGTGTGCGCATGGGGAAGTCGGAATTTCTTACTCCCAAGGCTA 91
Db 40 AGAAAGTGATCGCTGCGCGTGTGCGCATGGGGAAGTCGGAATTTCTTACTCCCAAGGCTA 99

Qy 92 TCGCAACAGAGATCAAGTCCAAAGGGGTGCAAGAGCTACGCTGGTATGTCAGATGTGCC 151
Db 100 TCGCAACAGAGATCAAGTCCAAAGGGGTGCAAGAGCTACGCTGGTATGTCAGATGTGCC 159

Qy 152 AGAGCAGTCCGGGAGAGAGATGCTTTAAGTGTGCTATGATGTCGGAATCTCATCAGA 211
Db 160 AGAGCAGTCCGGGAGAGAGATGCTTTAAGTGTGCTATGATGTCGGAATCTCATCAGA 219

Qy 212 GACAACTATTGCTGGCTTCAGAAATCCTCAGCAGTTCATGATTTATTTTCAGAGGAAT 271
Db 220 GACAACTATTGCTGGCTTCAGAAATCCTCAGCAGTTCATGATTTATTTTCAGAGGAAT 279

Qy 272 TCCGAATGATCTTTCTAGAACTTTCTCAGGAGACGCTTTGGCACTAAAGGGTCCACAACA 331
Db 280 TCCGAATGATCTTTCTAGAACTTTCTCAGGAGACGCTTTGGCACTAAAGGGTCCACAACA 339

Qy 332 ACATTGTCTAGACGAATACATCAGCCACGAGGACATCCATCAATGAATGCCACTCAGT 391
Db 340 ACATTGTCTAGACGAATACATCAGCCACGAGGACATCCATCAATGAATGCCACTCAGT 399

Qy 392 GGGAACTCTGACTGATTTTACTTAAGTGGCTGGGAGAGAGGCTTTGTGCAAA----- 444
Db 400 GGGAACTCTGACTGATTTTACTTAAGTGGCTGGGAGAGAGGCTTTGTGCAAAAGTGGACG 459

Qy 445 ----- 444
Db 460 AGACACCAAAAGGCTGGTATATTTCAGTACATAGACAGGAGCCAGAACTATCCGCGGC 519

Qy 445 ----- 444
Db 520 AACTGGAACTGGAGAAAAAGAAAAGCAGGACCTTTGATGATGAAGAAAAAAATGCAAAAT 579

Qy 445 ----- 444
Db 580 TTATTGAAGACGACGAGTGAAGAGCGCTTGGAAAGGAAGCAAGGAGGTCCTTACTTTTA 639

```

```

Qy 445 ----- 444
Db 640 CGGAATTAACGACAGAAAAATGATGAAGAGAAAGTCACGTTTAATTTGAGTAAAGGAGCAT 699

Qy 445 ----- 444
Db 700 GTAGCTCATCCGGAGCAACATCTTCCAAGTCAAGTACTCTGGGAGCCGAGTGCACGTGAAGA 759

Qy 458 CGATAGGAGTTTCAGCATCAGTCAGAACAAAAGAAATCTTCCAGAGCTCAACTCAGTCTA 517
Db 760 CGATAGGAGTTTCAGCATCAGTCAGAACAAAAGAAATCTTCCAGAGCTCAACTCAGTCTA 819

Qy 518 AAGAAAAGAGAAAAAGAAATCTGCACTGGATGAAATCATGGAGATTGAAGAGAGAAAGA 577
Db 820 AAGAAAAGAGAAAAAGAAATCTGCACTGGATGAAATCATGGAGATTGAAGAGAGAAAGA 879

Qy 578 AAGAAACTGCCGAGCAGACTACTGGCTACAGCTGAAATTTATGTAATAATTTATATACCA 637
Db 880 AAGAACTGCCGAGCAGACTACTGGCTACAGCTGAAATTTATGTAATAATTTATATACCA 939

Qy 638 AGAACTGGGAGAGAAATATCATAGAAAGAGGCTATTGTTAAGGAGTAGTAATTGACAAAT 697
Db 940 AGAACTGGGAGAGAAATATCATAGAAAGAGGCTATTGTTAAGGAGTAGTAATTGACAAAT 999

Qy 698 ATACAGCTGTTGTGAAGATGATTGATTTCTGGAGACAAGCTGAAAATTTGACAGACTCAT 757
Db 1000 ATACAGCTGTTGTGAAGATGATTGATTTCTGGAGACAAGCTGAAAATTTGACAGACTCAT 1059

Qy 758 TAGAGACAGTAATTCACGACACAGGAAAAAGAAATTTCTAGTTTTTAAATGGAGGCTACAGAG 817
Db 1060 TAGAGACAGTAATTCACGACACAGGAAAAAGAAATTTCTAGTTTTTAAATGGAGGCTACAGAG 1119

Qy 818 GAAATGAGTACCTTAGAATCCATCAATCAGAGAGACTTTTTCAGCTACTATCGCTATTG 877
Db 1120 GAAATGAGTACCTTAGAATCCATCAATCAGAGAGACTTTTTCAGCTACTATCGCTATTG 1179

Qy 878 AAATGCGCCCTTTTAAAGGACGAGAGTTTCAAGGAATTCATATGAAGAGACATTTCTAAAC 937
Db 1180 AAATGCGCCCTTTTAAAGGACGAGAGTTTCAAGGAATTCATATGAAGAGACATTTCTAAAC 1239

Qy 938 TTGCTGAGTTGAAAAATTTGTTAAACAATACCTTTAAAAATCTTAAAGCATCAAAATTTGGTG 997
Db 1240 TTGCTGAGTTGAAAAATTTGTTAAACAATACCTTTAAAAATCTTAAAGCATCAAAATTTGGTG 1298

Qy 998 TTGCGC 1002
Db 1299 TTGCGC 1303

RESULT 4
HSAJ5273
LOCUS HSAJ5273 1528 bp mRNA linear PRI 13-DEC-2000
DEFINITION Homo sapiens mRNA for Kin17 protein.
ACCESSION AJ005273
VERSION AJ005273.1 GI:3850703
KEYWORDS Kin17 gene.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1
AUTHORS Kannouche, P., Mauffrey, P., Pinon-Lataillade, G., Mattei, M. G.,
Sarasin, A., Daya-Grosjean, L. and Angulo, J. P.
TITLE Molecular cloning and characterization of the human KIN17 cDNA
encoding a component of the UVC response that is conserved among
metazoans
JOURNAL Carcinogenesis 21 (9), 1701-1710 (2000)
PUBMED 10964102
REFERENCE 2
AUTHORS Mauffrey, P.
TITLE Direct Submission

```

JOURNAL	Submitted (08-APR-1998) Mauffrey P., LGR, CEA-DSV-DRR, BP 6 - Fontenay aux Roses Cedex, 92265, FRANCE
REMARK	Revised by [3]
REFERENCE	3 (bases 1 to 1528)
AUTHORS	Mauffrey P.
TITLE	Direct Submission
JOURNAL	Submitted (07-OCT-1998) Mauffrey P., LGR, CEA-DSV-DRR, BP 6 - Fontenay aux Roses Cedex, 92265, FRANCE
COMMENT	On Nov 7, 1998 this sequence version replaced gi:3046739. Kannoche P., Mauffrey P., Pinon-Lataillade G., Biard D.S.F., Angulo J.F.
Sequences codant pour la Proteine kin17 et leurs Applications - French Patent Nr 97 15536 1:1-50(1997).	
FEATURES	Location/Qualifiers
	1..1528
source	/organism="Homo sapiens"
	/mol_type="mRNA"
gene	/db_xref="taxon:9606"
	/sex="male"
CDS	/tissue_type="testis"
	1..1528
	/gene="kin17"
	66..1247
	/genes="kin17"
	/codon_start=1
	/protein_id="CAA06462.1"
	/db_xref="GI:3850704"
	/db_xref="GOA:O60870"
	/db_xref="UniProt/TREMBL:O60870"
	/translations="MGKSDFLTPKAIANRISKSLGKLQKLRWYCMQCKQCRDENGFKCH CMSEHQRIQLLASENQPMDFPSEFRNDLELLRRFGTKRVHNNIYVNYBIYSHR EHHMNAOWETITDFPKWLGREGCLKVDETPKGWYIQYIDRDPETIRROLEKKKK QDLIDREKTAFTIEQVRRGLEKEQEVPTFELSRNDEKVTENLKGACSSGAT SXSXSTLGPAKLTIGSSAVKRSQSSQTSQSEKKKKXALDEIMEIEEKKRTAT TDWLQPEIIVKIITIKGSKYHKKAIVREIDKYTAVVMIMDSGDKLKDQTHLET VIPAPGKRRIILVNGYRGNEGTLSEINKEKTFSTATVIETGLKGRRVEGIQYEDISKLA"
	A"
ORIGIN	
Query Match	65.2%; Score 553.4; DB 8; Length 1528;
Best Local Similarity	76.6%; Pred. No. 2e-144;
Matches 969; Conservative	0; Mismatches 1; Indels 295; Gaps 2;
Qy	32 AGAAGTGATCGTCGCGTGTGCCCATGGGAAAGTCGATTTTCTTACTCCCAAGGCTA 91
Db	40 AGAAGTGATCGTCGCGTGTGCCCATGGGAAAGTCGATTTTCTTACTCCCAAGGCTA 99
Qy	92 TCGCCAA CAGGATCAAGTCCAAGGGGCTGCAGAGCTACGCTGGTATTGCCAGATGTGCC 151
Db	100 TCGCCAA CAGGATCAAGTCCAAGGGGCTGCAGAGCTACGCTGGTATTGCCAGATGTGCC 159
Qy	152 AGAAGCAGTCCGGGAGCAGAAATGGCTTTAAGTGTCAATTGATGTCGGAATCTCATCAGA 211
Db	160 AGAAGCAGTCCGGGAGCAGAAATGGCTTTAAGTGTCAATTGATGTCGGAATCTCATCAGA 219
Qy	212 GACAACTATTGCTGGCTTCAGAAAATCCTCAGCAGTATTATGGATTATTTTTCAGAGGAAT 271
Db	220 GACAACTATTGCTGGCTTCAGAAAATCCTCAGCAGTATTATGGATTATTTTTCAGAGGAAT 279
Qy	272 TCCGAAATGACTTTTCTAGAACTTCTCAGGAGACCGCTTTGGCAGCTATAAAGGGTCCACAACA 331
Db	280 TCCGAAATGACTTTTCTAGAACTTCTCAGGAGACCGCTTTGGCAGCTATAAAGGGTCCACAACA 339
Qy	332 ACATTGCTTCAACGAATATACATCAGCCA CCGAGAGCACAATCCACATGAATGCCACTCAGT 391
Db	340 ACATTGCTTCAACGAATATACATCAGCCA CCGAGAGCACAATCCACATGAATGCCACTCAGT 399
Qy	392 GGGAAACTCTGACTGATTTTACTAAGTGGCTGGGAGAGGCTCTGTGAAA----- 444
Db	400 GGGAAACTCTGACTGATTTTACTAAGTGGCTGGGAGAGGCTCTGTGAAAATGGAGCG 459
Qy	445 ----- 444

AUTHORS

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L.H., Shermen, C.M., Schuler, G.D., Altschul, S.F., Zebberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalios, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.

TITLE
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

PUBMED
12477932

REFERENCE
2 (bases 1 to 2069)

AUTHORS
Strausberg, R.

TITLE
Direct Submission

JOURNAL
Submitted (05-NOV-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK
NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT
On Dec 19, 2003 this sequence version replaced gi:16878214.

Contact: MGC help desk

Email: cgaps-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC),

Gaithersburg, Maryland;

Web site: <http://www.nisc.nih.gov/>

Contact: nisc_mgc@nhgri.nih.gov

Akhter, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Taurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H., and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 39 Row: O Column: 18
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 13124882.

FEATURES
source

1. 2069
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:29595 IMAGE:5089578"
/tissue_type="pancreas, epithelioid carcinoma"
/clone_lib="NIH MGC 42"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
1. 2069

gene

/gene="KIN"
/note="synonyms: KIN17, BTCD"
/db_xref="GeneID:22944"
/db_xref="MIM:601720"
12..1193

CDS

/gene="KIN"
/codon_start=1

/product="HsKin17 protein"
/protein_id="AAH17309.1"
/db_xref="GI:16878215"
/db_xref="GeneID:22944"
/db_xref="MIM:601720"
/translation="MGKSGFLTPKAIANRIKSKGLQKLWYCMQCKQCRDENGPKCH
CMSEHQQLLASNPQDMDFSEEFNDFLELRRFGRKRVHNNIVYVEIYSHR
EHIHNATQWETLDTFTKWLREGCLCKVDTPKGYIYIDRDPETIRGLKLEKKK
QDLDBEKTAKTIEERVRGLEGKEQVFTFTELRENDEKVTNLSKAGSSGAT
SSKSTLPSALKTIGSSASVRKSSOSSTQSEKKKKKSDALDEMETIEEKKRTAR
TDYWLQPEIIVKLIITKKLGEKHKKAIKVEIDKYTAVVMIDSDGLKLDQTHLET
VTPAPGKRLLVLNGGTGRNGEGLNESINEKTFATVIETGPKLGRVEGIQVEDISK
A"

ORIGIN

Query Match	63.7%	Score 638.4	DB 8	Length 2069
Best Local Similarity	76.3%	Pred. No. 7.3e-141		
Matches 954	Conservative 0	Mismatches 1	Indels 295	Gaps 2

QY	47	CCGTGTCGCCATGGGAAAGTCGGATTTTCTTACTCCCAAGGCTATCGCCAAACAGGATCA	106
DB	1	CCGTGTCGCCATGGGAAAGTCGGATTTTCTTACTCCCAAGGCTATCGCCAAACAGGATCA	60
QY	107	AGTCCAAAGGGCTGCAGAAAGCTACGCTGGTATTCAGATGTCGCCAGACAGCTGCGGG	166
DB	61	AGTCCAAAGGGCTGCAGAAAGCTACGCTGGTATTCAGATGTCGCCAGACAGCTGCGGG	120
QY	167	ACGAGAAATGGCTTTTAAGTGTCAATGTCGGAATCTCATCAGAGACAACTATTGCTGG	226
DB	121	ACGAGAAATGGCTTTTAAGTGTCAATGTCGGAATCTCATCAGAGACAACTATTGCTGG	180
QY	227	CTTCAGAAATCTCAGACAGTATTTATGGATTTATTTTTCAGAGGAATTCGAAATGACTTTC	286
DB	181	CTTCAGAAATCTCAGACAGTATTTATGGATTTATTTTTCAGAGGAATTCGAAATGACTTTC	240
QY	287	TAGAACTTCTCAGAGACGCTTTGGCACTAAAGGGTCCCAACAACTTGTCTACACG	346
DB	241	TAGAACTTCTCAGAGACGCTTTGGCACTAAAGGGTCCCAACAACTTGTCTACACG	300
QY	347	AATACATCAGCCACCGAGACATCCACATGAATGCCACTCAGTGGGAACTCTGACTG	406
DB	301	AATACATCAGCCACCGAGACATCCACATGAATGCCACTCAGTGGGAACTCTGACTG	360
QY	407	ATTTTACTAAGTGGCTGGGAGAGAGGCTTTGTGCAAA	444
DB	361	ATTTTACTAAGTGGCTGGGAGAGAGGCTTTGTGCAAAAGTGGAGACACCAAAAGGCT	420
QY	445	-----	444
DB	421	GGTATATTCACTACATAGACAGGAGCCAGAAACTATCCCGCGGCACTGGAACTGGAGA	480
QY	445	-----	444
DB	481	AAAAAGAAAGCAGGACCTTTGATGATGAAGAAAAAACTGCCAAATTTATTGAAGAGCAAG	540
QY	445	-----	444
DB	541	TGAGACAGGCGCTGGAAGGGAAGGACAGGAGGTCCTACTTTTACGGAATTAAGCAGAG	600
QY	445	-----	444
DB	601	AAATGATGAAGAGAAAGTCAACGTTTAAATTTAGTAAGAGGAGCATGTAGCTCATCCGGAG	660
QY	445	-----	472
DB	661	CAACATCTTCAAGTCAAGTACTCTCTGGGACCGAGTGCATCTGAGACCATAGGAAGTTTCAG	720
QY	473	CATCAGTGAACGAAAAAGAAATCTTCCAGAGCTCAACTCAGTCTTAAAGAAAGAGAAAA	532
DB	721	CATCAGTGAACGAAAAAGAAATCTTCCAGAGCTCAACTCAGTCTTAAAGAAAGAGAAAA	780
QY	533	AGAAATCTGCACTGGATGAATCATGGAGATTGAAGAGGAAAAAGAAAGAACTCCCGAA	592

Db	781	AGAAATCTGCACCTGGATGAAATCATCGAGATTCAGAGGAGAAAGAAAAGAACTGCCCGAA	840
QY	593	CAGACTACTGGCTACAGCCTGAAATTTATTGTGAAAAATTATAACCAAGAAACTGGGGAGAG	652
Db	841	CAGACTACTGGCTACAGCCTGAAATTTATTGTGAAAAATTATAACCAAGAAACTGGGGAGAG	900
QY	653	TATATCATAGAAAAGGCTATTGTTAAGGAAGTAATTGACAAATATACAGCTGTTGTGA	712
Db	901	AATATCATAGAAAAGGCTATTGTTAAGGAAGTAATTGACAAATATACAGCTGTTGTGA	960
QY	713	AGATGATTGATTCGGAGACAAGCTGAAACTTGACCAGACTCATTTAGAGACAGTAATTC	772
Db	961	AGATGATTGATTCGGAGACAAGCTGAAACTTGACCAGACTCATTTAGAGACAGTAATTC	1020
QY	773	CAGCACCAGAAAAAGAAATCTTAGTTTTAAATGGAGGCTACAGAGAAATGAAGGTACCC	832
Db	1021	CAGCACCAGAAAAAGAAATCTTAGTTTTAAATGGAGGCTACAGAGAAATGAAGGTACCC	1080
QY	833	TAGATCCATCATGAGAGACTTTTTTCAGCTACTATCGTCATTTGAACTGGCCCTTTAA	892
Db	1081	TAGATCCATCATGAGAGACTTTTTTCAGCTACTATCGTCATTTGAACTGGCCCTTTAA	1140
QY	893	AAGGACGACAGAGTTGAAGGAATTCATATGAAGACATTTCTAACTTGCTGAGTTTGAA	952
Db	1141	AAGGACGACAGAGTTGAAGGAATTCATATGAAGACATTTCTAACTTGCTGAGTTTGAA	1200
QY	953	AATTTGTTAAACATACCTTTTAAATCTTAAAGCATCAAAATGGTGTCCG	1002
Db	1201	AATTTGTTAAACATA-CAUTAAATCTTAAAGCATCAAAATGGTGTCCG	1249
RESULT 6			
AX003311			
LOCUS	AX003311	1002 bp	DNA linear PAT 24-AUG-2000
DEFINITION	Sequence 4 from Patent WO929845.		
ACCESSION	AX003311		
VERSION	AX003311.1	GI:9927128	
KEYWORDS	.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	1		
AUTHORS	Angulo-Mora, J.F. and Mauffrey, P.		
TITLE	Sequences coding for kin17 protein and their applications		
JOURNAL	Patent: WO 9929845-A 4 17-JUN-1999;		
FEATURES	Angulo MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE (FR)		
source	Location/Qualifiers		
	1. .1002		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
ORIGIN			
Query Match	46.9%;	Score 470;	DB 6; Length 1002;
Best Local Similarity	100.0%;	Pred. No. 8.1e-101;	
Matches	470;	Conservative 0;	Mismatches 0; Indels 0; Gaps 0;
QY	445	AGTGCACTGAAGCAGTAGGAGTTTCAGCATCAGTGAACCGAAAGAAATCTTCCAGAGC	504
Db	533	AGTGCACTGAAGCAGTAGGAGTTTCAGCATCAGTGAACCGAAAGAAATCTTCCAGAGC	592
QY	505	TCAACTCAGTCTAAAGAAAAGAGAAATCTGCATGGATGAAATCATGGAGATT	564
Db	593	TCAACTCAGTCTAAAGAAAAGAGAAATCTGCATGGATGAAATCATGGAGATT	652
QY	565	GAAGAGAAAAGAAAAGAACTGCCGACAGACTACTGGCTACAGCCTGAAATTTATTGTG	624
Db	653	GAAGAGAAAAGAAAAGAACTGCCGACAGACTACTGGCTACAGCCTGAAATTTATTGTG	712
QY	625	AAAAATTATACCAAGAAACTGGGAGAGAAATATATAGAAAAGGCTATTGTTAAGGAA	684
Db	945	AAAAATTATACCAAGAAACTGGGAGAGAAATATATAGAAAAGGCTATTGTTAAGGAA	1000

101 GGATCAAGTCAAGGGCTGCAGAGCTACGCTGGTATTGCCAGATGTCGCAGAGCAGT 160
 102 |||||
 63 GAATTAAGTCCAAAGGGCTCCAGAGCTTCCTGGTACTGCCAGATGTCGCAAGCAAT 122
 104 |||||
 161 GCGGGACGAGAATGGCTTTAAGTGTCAATGTATGTCGGAATCTCATCAGAGACAATAT 220
 106 |||||
 123 GCGCGACGAGAATGGCTTTAAGTGTCACTGTATGTCGAATCTCATCAAGAACAATGT 182
 108 |||||
 221 TGCTGGCTTCAGAAAATCCTCAGAGTTCATGATTTATTTTTCAGAGGAATTCGGAATG 280
 110 |||||
 183 TGCTGGCTTCAGAAAATCCTCAGAGTTCATGATTTATTTTTCAGAGGAATTCGGAATG 242
 112 |||||
 281 ACTTTCTAGAACTTCTCAGAGAGCTTTCGCACTAAAGGGTCCACAAACAATGTCT 340
 114 |||||
 243 ACTTTCTGGAATCTCTCAGGCGAGCTTTGGCACTAAAGGGTCCACAAACAATGTCT 302
 116 |||||
 341 ACAACGAATACATCAGCCACCGAGAGCACATCCACATGAATGCCACTCAGTGGGAAATC 400
 118 |||||
 303 ACAATGAATACATCAGCCACCGAGAGCACATCCACATGAAGCTTACTCAGTGGGAGAC 362
 120 |||||
 401 TGACTGATTTTAACTAAGTGGCTGGGACAGAGAGGCTTTGTGCAA----- 443
 122 |||||
 363 TGACCGACTTTTACCAAGTGGCTGGGACAGAGAGGCTTTGTGTAAGTGGATGAGACACGA 422
 124 |||||
 444 ----- 443
 423 AAGCTGGTAACTTCACTAGTACATAGACAGACACCAGAAAACCATCCGTCGGCACTGGAA 482
 126 |||||
 444 ----- 443
 483 TAGAAAAAAGAGACGACAGATCTGCAGCATGAAGAAAAAATGCCAAGTTCATTGAGG 542
 128 |||||
 444 ----- 443
 543 AGCAGGTGAGAGAGGCTCGAAGGAAAGAGCAGAGACACCTGTTTTCACAGAACTAA 602
 130 |||||
 444 ----- 443
 603 GCGAGAAAATGAGGAGAAAAGATTAGTTCAATTCGAATAAAGGAGCGGCTGGCTCAG 662
 132 |||||
 444 -----AAGTGCACTGAAGACGATAGGAA 466
 663 CGGAGGTACAAATCCAGTCAAGCTTTTGGGACCAAGTGCACTGAAGCTGTGGGA 722
 134 |||||
 467 GTTCAGCATCAGTCAAGAAAAGAAATCTTCCAGAGCTCAATCAGTCTAAAGAAAGA 526
 136 |||||
 723 GCGCAGCATCGGGAACGGAAGAGTCTTACAGAGCTCGCCGAGCCT-----GCGA 776
 138 |||||
 527 AGAAAAGAAATCTGCATCGATGAAATCATGGAGATTGAGAGGAAAGAAAGAACTG 586
 140 |||||
 777 AGAAGAGAAATCTGCCTCGATGAGATCATGGAGCTCGAAGAGGAAAGAAAGGACCG 836
 142 |||||
 587 CCCGAACAGACTACTGGCTACAGCTGAAATTTATGTAAGAAATTAATACCAAGAACTGG 646
 144 |||||
 837 CACGACAGAGCGCTGGTTACAGCGGGATCGTTGTGAAATTAATACGAGAGAGCTTG 896
 146 |||||
 647 GAGAGAAATATCATAAGAAAAGGCTATTTGTAAGGAAGTAAATGACAAATATACAGCTG 706
 148 |||||
 897 GGGAGAAATATCACAAGAAAGAAAGGGTCTTTAAGGAAGTATTGACAGGATACACAGCTG 956
 150 |||||
 707 TTGTGAAGATGATTTGATTTCTGAGACAGCTGAAATTTGACAGACTCATTTAGAGACAG 766
 152 |||||
 957 TGGTAAAGATGACTGACTCTGGAGACAGGCTGAAATCTGGACAGACTCATTTAGAGACAG 1016
 154 |||||
 767 TAATTCAGACACGAGAAAAGAAATTTCTAGTTTAAATGGAGGCTTACAGAGGAATGAAG 826
 156 |||||
 1017 TCATTCGCGCCCGGGGAAAGGGTTCTAGTTTAAATGGAGGCTTACAGAGGAATGAAG 1076
 158 |||||
 827 GTACCTTAGAATCCATCAATGAGAGACTTTTTCAGTACTATCGTCAATGAAACTGCC 886
 160 |||||
 1077 GCACCTCTCGATCCATCAATGAGAGGCTTTTTCAGCCACGATAGTCAATGAAACTGGAC 1136
 162 |||||
 887 CTTTAAAGGAGCGACAGTTGAAGGAATTCATATGAAAGACATTTCTTAAACTTGCCTGAG 946
 164 |||||

1137 CTTTGAAGGACGACAGATTTGAAGTATTCAATATGAAGACATATCTAAACTTGTCTGAG 1196
 1138 |||||
 947 TTTGAATAATTTGTAAACATACCTTTAAATCTTAAAGCATCAAAATTTGGTTCGC 1002
 1140 |||||
 1197 TTTGAATAATTTGTAAACACACATTTGAACT-GTGAAGCATCAAAATTTGATTTAGC 1251
 1142 |||||
 RESULT 10
 AC119716 221547 bp DNA linear HTG 19-NOV-2002
 LOCUS Rattus norvegicus clone CH230-274116, *** SEQUENCING IN PROGRESS
 DEFINITION *** 17 unordered pieces.
 AC119716
 AC119716.6 GI:25092366
 HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
 VERSION
 KEYWORDS Rattus norvegicus
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muroidae; Muridae; Murinae; Rattus.
 1 (bases 1 to 221547)
 Muzny, D., Marie, M., Metzker, M., Lee, S., Abramson, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Blawalo, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falle, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorensuhewa, L., Loulseghe, H., Lozano, R.J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
 Nwankweli, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plummer, R., Poindestre, A., Popovic, D., Primus, E., Pu, L.L.,
 Puzo, M., Quiroz, J., Rachin, E., Reeves, K., Regier, M.A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rodkey, T., Rojars, A., Rose, M., Rose, R., Ruiz, S.J.,
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sleson, I., Sitter, C.D., Smajs, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umanai, K.,
 Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P.,
 Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weinstock, G. and Gibbs, R.A.
 Direct Submission
 Unpublished
 TITLE
 JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 221547)

Worley,K.C.
Direct Submission
Submitted (01-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 221547)

REFERENCE
AUTHORS
TITLE
JOURNAL

Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Nov 19, 2002 this sequence version replaced gi:23611286.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GWGC
Center clone name: CH230-274116
----- Summary Statistics

Assembly program: Phrap; version 0.990329
Consensus quality: 208013 bases at least Q40
Consensus quality: 211685 bases at least Q30
Consensus quality: 213907 bases at least Q20
Estimated insert size: 199182; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

* 1 198201: contig of 198201 bp in length
* 198202 198301: gap of unknown length
* 198302 199749: contig of 1448 bp in length
* 199750 199849: gap of unknown length
* 199850 201115: contig of 1266 bp in length
* 201116 201215: gap of unknown length
* 201216 202338: contig of 1123 bp in length
* 202339 202439: gap of unknown length
* 202439 203500: contig of 1062 bp in length
* 203501 203600: gap of unknown length
* 203601 204905: contig of 1305 bp in length
* 204906 205005: gap of unknown length
* 205006 206193: contig of 1187 bp in length
* 206193 206292: gap of unknown length
* 206293 207385: contig of 1093 bp in length
* 207386 207486: gap of unknown length
* 207486 208519: contig of 1034 bp in length
* 208520 209759: gap of unknown length
* 209759 209858: gap of unknown length
* 209859 211441: contig of 1583 bp in length
* 211442 211541: gap of unknown length

* 211542 212876: contig of 1335 bp in length
* 212877 212976: gap of unknown length
* 212977 214356: contig of 1280 bp in length
* 214357 214356: gap of unknown length
* 214357 217003: contig of 2647 bp in length
* 217004 217103: gap of unknown length
* 217104 218459: contig of 1356 bp in length
* 218460 218559: gap of unknown length
* 218560 219788: contig of 1229 bp in length
* 219789 219888: gap of unknown length
* 219889 221547: contig of 1659 bp in length.

FEATURES
Location/Qualifiers

1..221547
source

/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-274116"

misc_feature

1..1272
/note="wgs end extension
clone_end:Sp6"
misc_feature
complement(4416..5263)
/note="clone boundary
clone_end:Sp6
site:
end_sequence:BZ266352"

misc_feature

192445..193553
/note="wgs_contig"

misc_feature

195743..196592
/note="clone boundary
clone_end:t7
site:
end_sequence:BZ266349"

gap

198202..198301
/estimated_length=unknown

gap

199750..199849
/estimated_length=unknown

gap

201116..201215
/estimated_length=unknown

gap

202339..202438
/estimated_length=unknown

gap

203501..203600
/estimated_length=unknown

gap

204906..205005
/estimated_length=unknown

gap

206193..206292
/estimated_length=unknown

gap

207386..207485
/estimated_length=unknown

gap

208520..208619
/estimated_length=unknown

gap

209759..209858
/estimated_length=unknown

gap

211442..211541
/estimated_length=unknown

gap

212877..212976
/estimated_length=unknown

gap

214257..214356
/estimated_length=unknown

gap

217004..217103
/estimated_length=unknown

gap

218460..218559
/estimated_length=unknown

gap

219789..219888
/estimated_length=unknown

Query Match 42.2%; Score 422.4; DB 14; Length 221547;

Best Local Similarity 66.0%; Pred No. 1.2e-89;

Matches 829; Conservative 0; Mismatches 126; Indels 301; Gaps 3;

QY

41 TCGCTGCCGTGCGCCATGGGAAAGTCGATTCTTACTCCCAAGGCTATGCCCAACA 100
|||||

DB

212995 TAGCGCTAGGCGTCGCATGGGCAAGTCGATTCTTCTGAGCCCAAGCCATGCCCAATA 213054
|||||

```
QY 101 GGATCAAGTCCAGGGGCTGCAGAGCTAGCTGGTATTGCCAGATGTGCCAGAACGAT 160
Db 213055 GAATTAAAGTCCAAAGGGCTCCAGAGCTTCGCTGGTACTGCCAGATGTGCCAAAGCAAT 213114
QY 161 GCCGGACGAGAAATGGCTTTAAAGTGTCAATGTATGTCCGAATCTCATCAGAGACAATCTAT 220
Db 213115 GCCGCGACGAGAAATGGCTTTAAAGTGTCACTGTATGTCTGAATCTCATCAAGAGACAATCTGT 213174
QY 221 TGTGCTGCTCAGAAATCTCAGCAGTATTATGATTTATTTTCAGAGGAATTCGGAATG 280
Db 213175 TGTGCTGCTCAGAAATCCCTCAGCAGTATTATGATTTATTTTCAGAGGAATTCGGAATG 213234
QY 281 ACTTTCTAGAACTTCTCAGAGAGCTTTGGCAGCTAAAGGGTCCACAACAATGTGTCT 340
Db 213235 ACTTTCTGGAACTTCTGAGCGAGCTTTGGCAGCTAAAGGGTCCACAACAATGTGTCT 213294
QY 341 ACAACGAATATCATCAGCCACCGAGAGCACATCCACATGAATGCCATCTCAGTGGGAATCTC 400
Db 213295 ACAATGAATATCATCAGCCACCGAGAGCACATCCACATGAACGCTACTCAGTGGGAGACAC 213354
QY 401 TGACTGATTTTACTAAGTGTGGCGAGAGAGGCTTGTGCAA----- 443
Db 213355 TGACCGACTTTTACCAAGTGGCTGGCAGAGAGGGCTTGTGTAAGTGGATGAGACACCGA 213414
QY 444 ----- 443
Db 213415 AAGCTGGTACATTCAGTACATAGACAGAGACCCAGAAACCCTCCGTCGGCAACTGGAT 213474
QY 444 ----- 443
Db 213475 TAGAAAAAAGNAGNACAGATCTGGACGATGAAGAAAAAACTGCCAAGTTCATTGAGG 213534
QY 444 ----- 443
Db 213535 AGCAGGTGAGAAGAGGCTCGAAGGGAAAGAGCAGCAGAGACACCTGTTTTTACAGAACTAA 213594
QY 444 ----- 443
Db 213595 GCCGAGAAATGAGGAAGAAAAAGTTACGTTCAATCTGAATTAAGGAGCGGGTGGCTCAG 213654
QY 444 -----AAGTGCACTGAAGACGATAGGAA 466
Db 213655 CGGGAGCTACAATCAATCAAGTCAAGCTCTTTGGGACCAAGTGCACTGAAGCTGTGGGA 213714
QY 467 GTTCAGCATCAGTGAAGAGAAAGAAATCTTCCAGAGCTCAACTCAGTCTTAAGAAAGA 526
Db 213715 GCGCAGCATCCGGGAAACGGAAGAGTCTTTCAGAGAGCTCGGCCCGCT-----CGCA 213768
QY 527 AGAAAAAGAAATCTGCACTGGTGAATCATGGAGATTGAAGAGGAAAGAAAGAACTG 586
Db 213769 AGAAGAAAGATGCGCCCTGGATGAGATCATGGAGCTCGAAGAGGAAAGAAAGAAAGACCG 213828
QY 587 CCCGAACAGACTACTGCTACAGCTGAAATTTATTTGTGAAAAATTAACCAAGAAACTGG 646
Db 213829 CACGACAGACGCTGTTACAGCGGGGATCGTTGTGAAAAATTAACGAGAACTTG 213888
QY 647 GAGAGAAATATCATPAGAAAAAGGCTATTGTTAAGGAAGTAATTGAATAATACAGCTG 706
Db 213889 GGGAGAAATATCATCAAGAAAGAAAGGGGCTGTTAAGGAAGTGAATGACAGGTACACAGCTG 213948
QY 707 TTGTGAAGATGATTGATCTGGAGACAAGCTGAACTTGACCACTGACCACTCATTTAGAGACAG 766
Db 213949 TGGTAAAGATGACTGACTCTGGAGACAAGGCTGAACTGGACCACTCATTTAGAGACAG 214008
QY 767 TAATTCAGCACACGAGAAAAAGAAATCTAGTTTTTAATTTGAAGGCTACAGAGGAATGAAG 826
Db 214009 TCATTCGGCCCCGGGGAAGAGGTTCTAGTTTTTAATTTGAAGGCTACAGAGGAATGAAG 214068
QY 827 GTACCCTAGAATCCATCAATGAGAGACTTTTTTCAGCTACTATCGTCAATGAAACTGGCC 886
Db 214069 GCACCTCGAATCCATCAATGAGAGGCTTTTTTCAGCCACGATAGTCAATGAAACTGGAC 214128
QY 887 CTTTAAAGGACGCGAGAGTTGAGGAATTCATATGAAGACATTTCTTAACCTTGCCTGAG 946
```

```
Db 214129 CTTTGAAGAGCGCAGAGTTGAAGGTATTCAATATGAAGACATATCTAAACTTGTCTGAG 214188
QY 947 TTTGAAAAATTTCTTAACATACCTTTTAAATCTTAAAGCATCAAATTTGTTGCTCGC 1002
Db 214189 TTTGAAAAATTTGATAACAACACATTTGAAACT-GTGAAGCATCAAATTTGATTTAGC 214243

RESULT 11
AX003331 1390 bp DNA linear PAT 24-AUG-2000
LOCUS Sequence 24 from Patent WO9929845.
DEFINITION AX003331
ACCESSION AX003331
VERSION AX003331.1 GI:9927146
KEYWORDS Mus sp.
SOURCE Mus sp.
ORGANISM Mus sp.
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
  Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
  Angulo-Mora, J.P. and Mauffrey, P.
  Sequences coding for kin17 protein and their applications
  Patent: WO 9929845-A 24 17-JUN-1999;
  ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE
  (FR)
FEATURES             Location/Qualifiers
     source            1..1390
                        /organism="Mus sp."
                        /mol_type="unassigned DNA"
                        /db_xref="taxon:10095"
ORIGIN
Query Match          41.6%; Score 416.6; DB 6; Length 1390;
Best Local Similarity 66.1%; Pred. No. 3.9e-88;
Matches 819; Conservative 0; Mismatches 119; Indels 301; Gaps 3;

QY 58 ATGGGAAAGTCGGATTTTCTTACTCCCAAGGCTATGCCCAAGGATCCCAACAGGATCAAGTCCAAGGG 117
Db 1 ATGGGCAAGTCGGATTTTCTGAGCCCAAGGCCATGCCCAATAGATTAAAGTCCAAGGG 60
QY 118 CTCAGAAAGCTACGCTGGTATTGCCAGATGTGCCAGAGCAGTGCCTGGGAGCAGAAATGGC 177
Db 61 CTCAGAAAGCTTCGCTGGTACTGCCAGATGTGCCAAAGCAATGCCGCGACGAAATGGC 120
QY 178 TTTAAGTGTCAATGTATGTCGGAATCTCATCAGACACACTATTGCTGGCTTCAGAAAT 237
Db 121 TTTAAGTGTCACTGTATGTCTGAATCTCATCAAAGACAACCTGTGTGGCTTCAGAAAC 180
QY 238 CCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTCGAAATGACTTTCTAGAACTTCTC 297
Db 181 CCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTCGAAATGACTTTCTGGAACCTCTG 240
QY 298 AGGAGAGCGCTTTGGCACTAAAAGGCTCCACAACAATTCTCTACAACGAATACATCAGC 357
Db 241 AGCGAGCGCTTTGGCACTAAAAGGCTCCACAACAATTCTCTACAATGAATAATCAGC 300
QY 358 CACCGAGAGCACATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGCTTTTACTAAG 417
Db 301 CACCGAGAGCACATCCACATGAATGCCACTCAGTGGGAGACACTGACCGACTTTTACCAG 360
QY 418 TGGCTGGGACAGAGAGGCTTTGTGCAA----- 443
Db 361 TGGCTGGGACAGAGAGGCTTTGTGTAAGTGGATGAGACACCGAAAGGCTGTGATTCAG 420
QY 444 ----- 443
Db 421 TACATAGACAGAGACCCAGAAACCCTCCGTCGGCAACTGGAATTAGAAAAAGAGAAG 480
QY 444 ----- 443
Db 481 CAAGATCTGGACGATGAAGAAAAAACTGCCAAGTTCAATTGAGGAGCGGTGAGAAGGCG 540
```


QY	444	-----	443
Db	541	CTGGAAGGGAAGAGCAGGAGACACCTGTTTTCAGAACTTAGCCGAGAAATGAGGAA	600
QY	444	-----	443
Db	601	GAAAAAGTTACGTTCAATCTGAATAAAGGAGCGGTGGCTCAGCGGGAGCTACAACATCC	660
QY	444	-----	443
Db	661	AAAGTCAAGCTCTTTGGGACCAAGTGCATGAAGCTGCTGGGAGCGCAGCATCCCGGAAA	720
QY	484	CGAAGAAGATCTCCAGAGCTCAACTCAGTCTTAAGAAAAGAGAAAAGAAATCTGCA	543
Db	721	CGGAAAGAGTCTTCACAGAGCTCCGCCAGCCT-----GCGAAGAAGAAAGAGTCCGCC	774
QY	544	CTGGATGAATCATGAGATTTGAAGAGGAAAAAGAAAAGAACTCCCGAACAAGACTACTGG	603
Db	775	CTGGATGAGATCATGGAGCTCGNAGAGAGAAAGAAAAGGAGCCGACGAGACGCTGG	834
QY	604	CTACAGCTGAAATTTATTTGTGAAAATTAACAAGAAACTGGGAGAGAAAATATCATAG	663
Db	835	TTACAGCCGGGATCGTTGTGAAAATTAATAACGAAGAAGCTTGGGAGAAAATATCACAA	894
QY	664	AAAAAGGCTATTGTTAAGAGAGTAATTGACAAATATACAGCTGTTGTGAAGATGATTGAT	723
Db	895	AGAAAGGGTCTGTTAAGGAAGTGTGACAGGTACACAGCTGTGTGAAGATGACTGAC	954
QY	724	CTCGAGACAAGCTGAAACTTGCACAGACTCATTTAGAGACAGTAATTCAGCACCAGGA	783
Db	955	CTCGAGACAGCTGAACTGACACAGACTCATTTAGAGACAGTCAATTCGCCCGGGG	1014
QY	784	AAAAGAAATCTAGTTTAAATGAGGCTACAGAGAAATGAAGGTACCTAGAACATCCATC	843
Db	1015	AAAAGGTTCTAGTTTAAATGAGGCTACAGAGAAATGAAGGCACTCTCGAATCCATC	1074
QY	844	ATGAGAACACTTTTCACTACTATCGTCATTGAACTGGCCCTTTAAAGAGCGCAGA	903
Db	1075	AATGAGAGGCTTTTTCAGCCAGATAGTCATTGAAACTGGACCTTTGAAGAGCGCAGA	1134
QY	904	GTGGAAGGAATCAATATGAAGACATTTCTAACTTGCCTGAGTTTGAATAATTTGTTAAC	963
Db	1135	GTGGAAGGTATTCAATATGAGACATATCTAACTTGTGATTTGAAAATTTGATATAC	1194
QY	964	ATATCTTTAAAATCTTAAAGCATCAAAATGGTGTTCG	1002
Db	1195	AAACATTTGAAACT-GTGAAGCATCAAAATTTGGTTAGC	1232
RESULT 12	BC058169	2710 bp mRNA linear ROD 13-FEB-2004	
LOCUS	BC058169	Mus musculus antigenic determinant of rec-A protein, mRNA (cdna	
DEFINITION	BC058169	clone IMAGE:5026202), partial cds.	
ACCESSION	BC058169	BC058169.1 GI:37194909	
VERSION	BC058169.1	GI:37194909	
KEYWORDS	BC058169.1	GI:37194909	
SOURCE	BC058169.1	GI:37194909	
ORGANISM	BC058169.1	GI:37194909	
REFERENCE	BC058169.1	GI:37194909	
AUTHORS	BC058169.1	GI:37194909	
QY	436	TTGTGCAAAAGTGCCTAGAACGACATAGAAAGTTCCAGCATCAGTGAACGAAAAGATCT	495
Db	2075	TTGGGACCAAGTGCCTAGAAAGTGTGGGAGCGCAGCATCCGGAAAACGGAAGAGTCT	2134
QY	496	TCCGAGAGTCAACTCAGTCTAAAGAAAAGAAAAGAAATCTGCATGATGAAATC	555
Db	2135	TCACAGAGCTCGCCCAACGCTGC-----GAAGAAGAAGAAGTCCGCCCTGATGAGATC	2188
QY	556	ATGGAGATTGAAGAGCAAAAGAAAAGAACTGCCCAACAGACTACTGGCTACAGCTGAA	615

Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahney, J., Helton, E., Kettner, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krawinski, M.I., Skalek, U., Smalish, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 2710)

Strausberg, R.

Direct Submission

Submitted (15-SEP-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs@mail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNL)

DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland;

Web site: <http://www.nisc.nih.gov/>

Contact: nisc_mgc@hgri.nih.gov

Akter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Young, A., Zhang, L.-H., Walker, M.A., Wetherby, K.D., Wiggins, L., and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LILNL at: <http://image.llnl.gov>

Series: IRAP Plate: 123 Row: d Column: 16

This clone was selected for full length sequencing because it passed the following selection criteria: GenomeScan gene prediction, Similarity but not identity to protein.

Location/Qualifiers

1..2710

/organism="Mus musculus"

/mol_type="mRNA"

/strain="mix FVB/N, C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:5026202"

/tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months old, gross tissue."

/clone_lib="NCI CGAP_Mam5"

/lab_host="DH10B"

/note="Vector: pCMV-SPORT6"

Query Match 39.0%; Score 390.6; DB 9; Length 2710;

Best Local Similarity 83.1%; Pred No. 5, 7e-82;

Matches 471; Conservative 0; Mismatches 89; Indels 7; Gaps 2;

QY 436 TTGTGCAAAAGTGCCTAGAACGACATAGAAAGTTCCAGCATCAGTGAACGAAAAGATCT 495

Db 2075 TTGGGACCAAGTGCCTAGAAAGTGTGGGAGCGCAGCATCCGGAAAACGGAAGAGTCT 2134

QY 496 TCCGAGAGTCAACTCAGTCTAAAGAAAAGAAAAGAAATCTGCATGATGAAATC 555

Db 2135 TCACAGAGCTCGCCCAACGCTGC-----GAAGAAGAAGAAGTCCGCCCTGATGAGATC 2188

QY 556 ATGGAGATTGAAGAGCAAAAGAAAAGAACTGCCCAACAGACTACTGGCTACAGCTGAA 615


```

Db      2189  ATGGAGCTCGAAGGAGAAAGAAAGGACCGCAGCAGCAGCAGCCCTGGTTACAGCCGGGG 2248
Qy      616  ATTATTGTGAAAATTATAACCAAGAAACTCGGAGAGAGAAATATCATAGAAAAGGCTATT 675
Db      2249  ATCGTTGTGAAAATTATAACGAAGAAGCTTTGGGAGAGAAATATCACAGAAGAAAGGGGTC 2308
Qy      676  GTTAAAGGAAGTAAATGACAATATACAGCTGTGTGTGAAGATGATTGATCTCGAGACAAAG 735
Db      2309  GTTAAAGGAAGTGAATGACAGGTACACAGCTGTGTGTGAAGATGACTGACTCTCGAGACAGG 2368
Qy      736  CTGAAACTTCACACAGACTCAATTAGAGACAGTAATATCCAGCACCAGCAAGAAAAGAACTTA 795
Db      2369  CTGAACTCGACACAGACTCATTTAGAGACAGTCAATTCGCGCCCGGGGAAAGGGTTCTA 2428
Qy      796  GTTTTAAATGAGGCTACAGAGGAATGAAGGTACCTAGAAATCCATCATCATGAGAGACT 855
Db      2429  GTTTTAAATGAGGCTACAGAGGAATGAAGGCACTCTCGAATCCATCATGAGAGAGGCT 2488
Qy      856  TTTTCAGCTACTATCGTTCATTGAAACTGGCCCTTTAAAGGACGACAGAGTTGAAGGAATT 915
Db      2489  TTTTCAGCCACGATAGTCATTGAAACTGGACCTTTGAAGAGCAGCAGAGTTGAAGGTATT 2548
Qy      916  CAATATCAAGACATTTCTAAACTTCGCTGAGTTTGAAGAAATTTGTTAAACAATACCTTTAAA 975
Db      2549  CAATATCAAGACATATCTAACTTGTGATTTGAAATTTGATAACAACACATTGAAA 2608
Qy      976  ATCTTAAAGCATCAAAATGCTGTTCGC 1002
Db      2609  CT-GTGAAGCATCAAAATGCTGTTCGC 2634

RESULT 13
AC114615/c
LOCUS   AC114615
DEFINITION Mus musculus clone RP24-82119, WORKING DRAFT SEQUENCE, 4 ordered
pieces.
ACCESSION AC114615
VERSION   AC114615.5 Gi:44886278
HTG; HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.
KEYWORDS  Mus musculus (house mouse)
ORGANISM  Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidae; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 180882)
AUTHORS  Birren,B., Nusbaum,C. and Lander,E.
TITLE    Mus musculus, clone RP24-82119
JOURNAL  Unpublished
REFERENCE 2 (bases 1 to 180882)
AUTHORS  Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collamore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kanat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Lander,A., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
Mihova,T., Miens,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,J., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (10-MAR-2002) Whitehead Institute/MIT Center for Genome

```

REFERENCE
AUTHORS

Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 180882)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collamore,A., Cook,A., Cooke,P., Dorris,L., DeArellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kanat,A., Karatas,A., Kells,C., Lander,E., Levine,R.,
Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Mihova,T., Miens,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (09-MAR-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 2, 2004 this sequence version replaced gi:24182248.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L24315
Center clone name: 82_1_19
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 191018 bases at least Q40
Consensus quality: 191303 bases at least Q30
Consensus quality: 191388 bases at least Q20
Insert size: 178000; agarose-fp
Insert size: 191574; sum-of-contigs
Quality coverage: 12.2 in Q20 bases; agarose-fp
Quality coverage: 11.3 in Q20 bases; sum-of-contigs

TITLE
JOURNAL

COMMENT

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 89963: contig of 89963 bp in length
* 89964 90063: gap of 100 bp
* 90064 94560: contig of 4497 bp in length
* 94561 94660: gap of 100 bp
* 94661 169601: contig of 74941 bp in length
* 169602 169701: gap of 100 bp
* 169702 180882: contig of 1181 bp in length.
FEATURES
source
1. 180882
/organism="Mus musculus"
/molecule="genomic DNA"
/db_xref="taxon:10090"
/clone="RP24-82119"
/clone_lib="RPC1-24 Male Mouse BAC"
misc_feature 1. 89963

/note="assembly_fragment clone_end:SP6 vector_side:left" 89964..90063 /estimated_length=100 90064..94560 /note="assembly_fragment" 94561..94660 /estimated_length=100 94661..169601 /note="assembly_fragment" 169602..169701 /estimated_length=100 169702..180882 /note="assembly_fragment clone_end:T7 vector_side:right"		39.0%; Score 390.6; DB 14; Length 180882; Best Local Similarity 83.1%; Pred. No. 4.5e-82; Matches 471; Conservative 0; Mismatches 89; Indels 7; Gaps 2;	
gap	436 TTGTGCAAAAGTGCCTGAAGACGATAGGAAGTTTCAGCATTCAGTGAACGAAAGAAATCT	495	
misc_feature	46568 TTGGGACCAAGTGCATCAAGCTGTGGGAGCGCAGCATCCGGGAAACGGAAAGAGTCT	46509	
gap	496 TCCGAGAGCTCAACTCAGTCTAAAGAAAAAGAAAAAGAAATCTGCATCTGGATGAAATC	555	
misc_feature	46508 TCACAGAGCTCCGCCAGCTGC-----GAAGAAGAAGAGTCCGCCCTGGATGATC	46455	
gap	556 ATGGAGATTGAAGAGAAAAAGAAAGAACTCCCGAACAGACTATCTGGCTCAGACCTGAA	615	
misc_feature	46454 ATGGAGCTCGAAGAGAAAAAGGACCGCACGACAGACGCTGGTTACAGCCGGG	46395	
gap	616 ATTATTGTGAAATTTATACCAAGAACTGGGAGAGAAATATCATAAAGAAAGGCTATT	675	
misc_feature	46394 ATCGTTGTGAAATTTATACGAAGAAAGCTTGGGGAGAAATATCACAAGAAAGAGGGTCT	46335	
gap	676 GTTAAGGAAGTAAATTCACAAATATACAGCTGTTGTGAAGATGATTGATTTCTGGAGACAAG	735	
misc_feature	46334 GTTAAGGAAGTATTGACAGGTTACACAGCTGTGTTAAGATGACTGACTCTGGAGACAGG	46275	
gap	736 CTGAAACTTGACAGACTCATTTAGAGACAGTAATTTCCAGCACGACGAAAGAAAGAAATCTTA	795	
misc_feature	46274 CTGAAACTTGACAGACTCATTTAGAGACAGTCAATTCGCGCCCGCGGAAAGGGTTCTA	46215	
gap	796 GTTTTAAATGGAGGCTACAGAGAAATGAAGTACCCTAGATCCATCAATCAGAGAGACT	855	
misc_feature	46214 GTTTTAAATGGAGGCTACAGAGAAATGAAGGCATCTCGAATCCATCAATGAGAGGCT	46155	
gap	856 TTTTTCAGCTACTATCGTCAATGAAACTGGCCCTTTAAAAAGGACGACAGAGTTGAAAGGAATT	915	
misc_feature	46154 TTTTTCAGCCACGATAGTCAATGAACTGGACCTTTGAAGAGCGACAGAGTTGAGGTATT	46095	
gap	916 CAATATGAAGACATTTCTAAACTTGCTCAGTTTGAAATTTGTTAAACAATACCTTTAAA	975	
misc_feature	46094 CAATATGAAGACATATCTAAACTTGCTCAGTTTGAAATTTGATTAACAACAACATTGAA	46035	
gap	976 ATCTTAAAGCATCAAAATTTGGTGTTCGC	1002	
misc_feature	46034 CT-GTGAAGCATCAAAATTTGGTGTTCG	46009	
RESULT 14			
AC124010			
LOCUS	AC124010 182400 bp DNA linear HTG 25-AUG-2002		
DEFINITION	Mus musculus chromosome UNK clone RP23-119N4, WORKING DRAFT		
ACCESSION	SEQUENCE, 6 unordered pieces.		
VERSION	AC124010		
KEYWORDS	HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP.		
SOURCE	Mus musculus (house mouse)		

ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 182400)
TITLE	McPherson,J.D. and Waterston,R.H.
JOURNAL	The sequence of Mus musculus clone
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 182400)
TITLE	McPherson,J.D. and Waterston,R.H.
JOURNAL	Direct Submission
REFERENCE	Submitted (06-JUN-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
AUTHORS	3 (bases 1 to 182400)
TITLE	McPherson,J.D. and Waterston,R.H.
JOURNAL	Direct Submission
COMMENT	Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA On Aug 25, 2002 this sequence version replaced gi:21327639.
----- Genome Center -----	
Center: Washington University Genome Sequencing Center	
Center code: WUGSC	
Web site:http://genome.wustl.edu/gsc/index.shtml	
Contact: submissions@watson.wustl.edu	
----- Project Information -----	
Center project name: M_BA0119N04	
----- Summary Statistics -----	
Sequencing vector: M13; 0%	
Chemistry: Dye-primer ET; 0% of reads	
Assembly program: Phrap; version 0.990319	
Consensus quality: 183528 bases at least Q40	
Consensus quality: 184210 bases at least Q30	
Consensus quality: 184756 bases at least Q20	
Insert size: 197000; agarose-fp	
Quality coverage: 9.53 in Q20 bases; agarose-fp	
Quality coverage: 9.61 in Q20 bases; sum-of-contigs	
* NOTE: This is a 'working draft' sequence. It currently	
* consists of 6 contigs. The true order of the pieces	
* is not known and their order in this sequence record is	
* arbitrary. Gaps between the contigs are represented as	
* runs of N, but the exact sizes of the gaps are unknown.	
* This record will be updated with the finished sequence	
* as soon as it is available and the accession number will	
* be preserved.	
* 1 11868: contig of 11868 bp in length	
* 11869 11968: gap of unknown length	
* 28907 28906: contig of 16938 bp in length	
* 29007 29006: gap of unknown length	
* 72773 72773: contig of 43767 bp in length	
* 72774 72873: gap of unknown length	
* 124927 124926: contig of 52053 bp in length	
* 125027 125026: gap of unknown length	
* 181903 181902: contig of 56876 bp in length	
* 182003 182002: gap of unknown length	
* 182400 182400: contig of 398 bp in length.	
Location/Qualifiers	
1..182400	
/organism="Mus musculus"	
/mol_type="genomic DNA"	
/db_xref="taxon:10090"	
/chromosome="UNK"	
/clone="RP23-119N4"	
1..11868	
/note="assembly_name:Contig86"	
11869..11968	
/estimated_length=unknown	
11869..28906	
misc_feature	
gap	
misc_feature	

```
gap /note="assembly_name:Contig87"
28907..29006
/estimated_length=unknown
misc_feature /note="assembly_name:Contig88"
29007..72773
72774..72873
/estimated_length=unknown
gap /note="assembly_name:Contig89"
72874..124926
124927..125026
/estimated_length=unknown
misc_feature /note="assembly_name:Contig90"
125027..181902
181903..182002
/estimated_length=unknown
gap /note="assembly_name:Contig91"
182003..182400
182401..182500
/estimated_length=unknown
misc_feature /note="assembly_name:Contig96"
```

ORIGIN

```
Query Match 39.0%; Score 390.6; DB 14; Length 182400;
Best Local Similarity 83.1%; Pred. No. 4.5e-82;
Matches 471; Conservative 0; Mismatches 89; Indels 7; Gaps 2;

Qy 436 TTGTGCAAAAGTGCACCTGAAGACGATAGGAAGTTTCAGCATCAGTGAACGAAAGAAATCT 495
Db 168845 TTGGACCAAGTGCACCTGAAGTCTCTGGGAGCGCAGCATCCGGAAACGGAAGAGTCT 168904

Qy 496 TCCAGAGCTCACTCAGTCTTAAGAAAGAAAGAAAGAAATCTGCACCTGGATGAATC 555
Db 168905 TCACAGAGCTCCGCCAGCCTGC-----GAAGAAGAAAGAGTCCGCCCTGGATGAGATC 168958

Qy 556 ATGAGATTGAGAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 615
Db 168959 ATGAGAGCTCCGCCAGCCTGC-----GAAGAAGAAAGAGTCCGCCCTGGATGAGATC 169018

Qy 616 ATTATTGTGAAATTTATACCAAGAAATGAGGAGAGAAATATCATAGAAAGAAAGCTATT 675
Db 169019 ATCGTTGTGAAATTTATACCAAGAAATGAGGAGAGAAATATCATAGAAAGAAAGCTATT 169078

Qy 676 GTTAAGGAAGTAATTGACAAATATACAGCTCTTGTGAAGATGATGATTCCTGGAGACAAG 735
Db 169079 GTTAAGGAAGTAATTGACAGGTACACAGCTCTGTTGAAGATGATGATTCCTGGAGACAAG 169138

Qy 736 CTGAAACTTGACAGACTCATTTAGACAGATTAATTCAGCACCAGGAAAGAAAGAAATCTTA 795
Db 169139 CTGAAACTTGACAGACTCATTTAGACAGATTAATTCAGCACCAGGAAAGAAAGAAAGTCTTA 169198

Qy 796 GTTTTAAATGAGGCTACAGAGAAATGAAGTACCCTTAGAATCCATCAATGAGAAAGCT 855
Db 169199 GTTTTAAATGAGGCTACAGAGAAATGAAGTACCCTTAGAATCCATCAATGAGAAAGCT 169258

Qy 856 TTTTCAGCTACTATCGTCAATGAAGTCCCTTTAAAGGAGCAGAGTGTGAAGGAAT 915
Db 169259 TTTTCAGCTACTATCGTCAATGAAGTCCCTTTAAAGGAGCAGAGTGTGAAGGAAT 169318

Qy 916 CAATATGAAGCACTTTCTAAACTTCCTGAGTTTGAATTTGTAATTTGTAATTTGTAATTT 975
Db 169319 CAATATGAAGCACTTTCTAAACTTCCTGAGTTTGAATTTGTAATTTGTAATTTGTAATTT 169378

Qy 976 ATCTTAAGCATCAAAATTTGGTTCGC 1002
Db 169379 CT-GTGAAGCATCAAAATTTGGTTCGC 169404
```

```
RESULT 15
LOCUS AL772367/c 185378 bp DNA linear ROD 02-SEP-2002
DEFINITION Mouse DNA sequence from clone RP23-119N4 on chromosome 2, complete
sequence.
ACCESSION AL772367
VERSION AL772367.7 GI:22759512
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
```

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 185378)
Wood, J.
Direct Submission
Submitted (30-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Sep 9, 2002 this sequence version replaced gi:2265488.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk

COMMENT

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-119N4 is
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm

FEATURES
source

Location/Qualifiers
1..185378
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="2"
/clone_lib="RP23-119N4"
/clone_lib="RPCI-23"

ORIGIN

```
Query Match 39.0%; Score 390.6; DB 9; Length 185378;
Best Local Similarity 83.1%; Pred. No. 4.5e-82;
Matches 471; Conservative 0; Mismatches 89; Indels 7; Gaps 2;

Qy 436 TTGTGCAAAAGTGCACCTGAAGACGATAGGAAGTTTCAGCATCAGTGAACGAAAGAAATCT 495
Db 139982 TTGGACCAAGTGCACCTGAAGTCTCTGGGAGCGCAGCATCCGGAAACGGAAGAGTCT 139923

Qy 496 TCCAGAGCTCACTCAGTCTTAAGAAAGAAAGAAAGAAATCTGCACCTGGATGAATC 555
Db 139922 TCACAGAGCTCCGCCAGCCTGC-----GAAGAAGAAAGAGTCCGCCCTGGATGAGATC 139869

Qy 556 ATGAGATTGAGAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 615
Db 139868 ATGAGAGCTCCGCCAGCCTGC-----GAAGAAGAAAGAGTCCGCCCTGGATGAGATC 139809

Qy 616 ATTATTGTGAAATTTATACCAAGAAATGAGGAGAGAAATATCATAGAAAGAAAGCTATT 675
Db 139808 ATCGTTGTGAAATTTATACCAAGAAATGAGGAGAGAAATATCATAGAAAGAAAGCTATT 139749

Qy 676 GTTAAGGAAGTAATTGACAAATATACAGCTCTTGTGAAGATGATGATTCCTGGAGACAAG 735
Db 139748 GTTAAGGAAGTAATTGACAGGTACACAGCTCTGTTGAAGATGATGATTCCTGGAGACAAG 139689
```

QY	736	CTGNAACTTGACCAGACTCATTTAGAGACAGTAATTCAGCACCAGGAAAAAGAAATTTCTA	795
Db	139688		
QY	796	CTGAAACTGGACCAGACTCATTTAGAGACAGTCATTCGGCCCCCGGGGAAAAAGGGTTCTA	139629
Db	139628		
QY	856	GTGTTTAAATGGAGGCTACAGAGGAAATGAAGGTACCTAGAAATCCATCAATGAGAAGACT	855
Db	139568		
QY	916	GTGTTTAAATGGAGGCTACAGAGGAAATGAAGGCATCTCGMATCCATCAATGAGAAGGCT	139569
Db	139509		
QY	976	TTTTCAAGTACTATCGTCAATGAAACTGGCCCTTTAAAAAGGACGCAGAGTTGAAGGAATT	915
Db	139448		
QY	976	TTTTCAGCCACGATAGTCATTTGAAACTGGACCTTTCAAAGGACGCAGAGTTGAAGGTATT	139509
Db	139449		
QY	976	CAATATGAAGACATTTCTAAACTTGCCTGAGTTTGAAAAATTTGTTAAACAATACCTTTAAA	975
Db	139449		
QY	976	CAATATGAAGACATATCTAAACTTGCCTGAGTTTGAAAAATTTGATAACAACACATTTGAAA	139449
Db	139448		
QY	976	ATCTTAAAGCATCAAATTTGGTGTTCGC	1002
Db	139448		
QY	976	CT-GTGAAGCATCAAATTTGGTGTTAGC	139423
Db	139448		

Search completed: November 26, 2005, 19:10:59
Job time : 5335.7 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

QM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 12:48:57 ; Search time 642.701 Seconds
(without alignments)
10390.557 Million cell updates/sec

Title: US-09-555-529-3
Perfect score: 1002
Sequence: 1 tgattcagctggtaccgc.....agcatcaaatggtgttcgc 1002

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq_21:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*
14: Geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1002	100.0	1002	2	AAX85551 CDNA of a
2	704.6	70.3	1102	2	AAX85550 CDNA of a
3	696.4	69.5	1296	2	AAX85549 CDNA of a
4	470	46.9	1002	2	AAX85552 Probe der
5	426.4	42.6	679	13	ADQ56943 Novel car
6	416.6	41.6	1390	2	AAX85570 CDNA of a
7	415.2	41.4	1458	2	AAG79936 Murine Ki
8	267.2	26.7	591	4	AAB34095 Human col
9	201.4	20.1	1372	12	ADP22451 Sea-squi
10	176	17.6	1713	13	ADT17629 Plant CD
11	176	17.6	1780	3	AA52591 CDNA enco
12	176	17.6	1780	3	AA52590 CDNA enco
13	171	17.1	300	3	AA01559 Human col
14	170.8	17.0	1614	13	ADO84255 plant fu
15	170.8	17.0	1713	13	ADX47874 plant fu
16	170.8	17.0	1764	3	AA52589 CDNA enco
17	168.2	16.8	1459	3	AAC47248 Arabidops
18	163.4	16.3	1241	4	ABL29199 Drosophil
19	163.4	16.3	3313	4	ABL29198 Drosophil

20	163.4	16.3	4394	4	ABL29184
21	163.4	16.3	4582	4	ABL18194
22	136.4	13.6	1032	3	AAC48794
23	126.4	12.6	549	3	AAC94929
24	100.2	10.0	270	6	ABL73715
25	97.2	9.7	398	13	ACF87901
26	94.2	9.4	696	3	AAC38477
27	77.4	7.7	431	3	AAC48801
28	60	6.0	60	6	ABN33592
29	56.8	5.7	2985	5	ABL29210
30	54.4	5.4	65	6	ABN51894
31	50.6	5.0	537	13	ACN53339
32	50	5.0	778	6	ABQ15589
33	50	5.0	778	6	ABQ15589
34	48.2	4.8	2000	11	ACL35887
35	48	4.8	2000	8	ADA71938
36	47.4	4.7	5940	3	AA70105
37	46.2	4.6	9964	6	ABL32098
38	45.2	4.5	461	13	ACN55894
39	44.6	4.5	4316	10	ADB54224
40	44.6	4.5	4316	13	ADS89522
41	44.6	4.5	6182	6	ABL34015
42	44	4.4	6821	3	AA70188
43	43.8	4.4	16888	4	AAS46556
44	43.6	4.4	529	10	ACC42649
45	43.6	4.4	612	14	ACL64894

ALIGNMENTS

RESULT 1
AAX85551
ID AAX85551 standard; cDNA; 1002 BP.

XX AAX85551;

DT 07-SBP-1999 (first entry)

DE cDNA of a gene coding for a truncated human kin17 protein.

XX kin17 protein; cell proliferation; fertility; hyperproliferative disease;
XX protein interaction; curved DNA; HIV replication; HIV integration;
XX repair enzyme; ss.

XX Homo sapiens.

XX FR2772046-A1.

XX 11-JUN-1999.

XX 09-DEC-1997; 97FR-00015536.

XX 09-DEC-1997; 97FR-00015536.

XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;

XX WPI; 1999-359999/31.

XX New DNA coding for human kin17 protein - useful for controlling cell proliferation or fertility.

XX Claim 5; Page 32; 69pp; French.

XX The present sequence encodes a truncated human kin17 protein with amino acids 129-228 deleted. The mammalian kin17 protein is useful for preparing a medicament for controlling cell proliferation or for controlling fertility. The medicaments can also be used to treat hyperproliferative diseases. Fragments between amino acids 55 and 235 (preferably between amino acids 129 and 228) of a mammalian kin17 protein are useful for regulating the interaction between proteins and curved

CC DNA. The fragment can be used to block replication of HIV or its
CC integration into the human genome or to target repair enzymes to curved
CC DNA sites. Expression vectors for kin17 can be used for controlling cell
CC proliferation
XX
SQ Sequence 1002 BP; 343 A; 189 C; 230 G; 240 T; 0 U; 0 Other;

Query Match 100.0%; Score 1002; DB 2; Length 1002;
Best Local Similarity 100.0%; Pred. No. 1.5e-265;
Matches 1002; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGATTCGAGTCGGTACCGGGATCCGATTAGAAAAGTCATCGCTCGGTGGTCCGATG 60
DB 1 TGATTCGAGTCGGTACCGGGATCCGATTAGAAAAGTCATCGCTCGGTGGTCCGATG 60

QY 61 GGGAGTCGGATTTCTTACTCCCAAGGCTATGCCAACAGGATCAAGTCCCAAGGGGCTG 120
DB 61 GGGAGTCGGATTTCTTACTCCCAAGGCTATGCCAACAGGATCAAGTCCCAAGGGGCTG 120

QY 121 CAGAAGCTACGCTGATTTGCCAGATGTGCCAAGCAGTGCCTGGGACGAGAATGGCTTT 180
DB 121 CAGAAGCTACGCTGATTTGCCAGATGTGCCAAGCAGTGCCTGGGACGAGAATGGCTTT 180

QY 181 AAGTGTCAATGTATGTCGGAATCTCATCAGAGCAACTATTCGTGGCTTCAGAAAATCCT 240
DB 181 AAGTGTCAATGTATGTCGGAATCTCATCAGAGCAACTATTCGTGGCTTCAGAAAATCCT 240

QY 241 CAGCAGTTATGGATTAATTTTTCAGAGGAATTCGGAATGACITTTCTAGAACTTCTCAGG 300
DB 241 CAGCAGTTATGGATTAATTTTTCAGAGGAATTCGGAATGACITTTCTAGAACTTCTCAGG 300

QY 301 AGACGCTTTGGCACTAAAAGGGTCCACAAACATTTGCTACAAAGATACATCAGCCAC 360
DB 301 AGACGCTTTGGCACTAAAAGGGTCCACAAACATTTGCTACAAAGATACATCAGCCAC 360

QY 361 CGAGAGCACATCCACATGAATGCCACTCAGTGGGAACTCTGACTCATTTTACTAAGTGG 420
DB 361 CGAGAGCACATCCACATGAATGCCACTCAGTGGGAACTCTGACTCATTTTACTAAGTGG 420

QY 421 CTGGGCAGAGAAGGCTTTGTGCAAAAGTGCACCTGAAGACGATAGGAAGTTTCAGCATCAGTG 480
DB 421 CTGGGCAGAGAAGGCTTTGTGCAAAAGTGCACCTGAAGACGATAGGAAGTTTCAGCATCAGTG 480

QY 481 AAACGAAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAGAAAAGAAATCT 540
DB 481 AAACGAAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAGAAAAGAAATCT 540

QY 541 GCACCTGATGAATCATGTGAGATTGAAGAGGAAAGAAAGAACTGCCCGAACAGACTAC 600
DB 541 GCACCTGATGAATCATGTGAGATTGAAGAGGAAAGAAAGAAAGAACTGCCCGAACAGACTAC 600

QY 601 TGCTACAGCTGAAATTAATTTGTGAAAATTTATACCAAGAAACTGGGAGAGAAATATCAT 660
DB 601 TGCTACAGCTGAAATTAATTTGTGAAAATTTATACCAAGAAACTGGGAGAGAAATATCAT 660

QY 661 AGAAAAAGCTTAATTTGTAAGGAAGTAAATTGACAAATATACAGCTGTTGTGAAGATGAT 720
DB 661 AGAAAAAGCTTAATTTGTAAGGAAGTAAATTGACAAATATACAGCTGTTGTGAAGATGAT 720

QY 721 GATTCCTGGAGACAAGCTGAAAATTGACAGACTCATTTAGAGACAGTAATTCAGCACCA 780
DB 721 GATTCCTGGAGACAAGCTGAAAATTGACAGACTCATTTAGAGACAGTAATTCAGCACCA 780

QY 781 GGAAGAAGAAATTCAGTCTTTTAAATGGAGGCTACAGAGGAAATGAAGGTACCCCTAGAATCC 840
DB 781 GGAAGAAGAAATTCAGTCTTTTAAATGGAGGCTACAGAGGAAATGAAGGTACCCCTAGAATCC 840

QY 841 ATCAATGAGAAGACTTTTTCAGCTACTATTCGTAATGAACTGGCCCTTTAAAGACGC 900
DB 841 ATCAATGAGAAGACTTTTTCAGCTACTATTCGTAATGAACTGGCCCTTTAAAGACGC 900

QY 901 AGAGTTTGAAGGAATTCAAATATGAAGACATTTCTTAAACTTCCTGAGTTTGAAATTTGTT 960
DB 901 AGAGTTTGAAGGAATTCAAATATGAAGACATTTCTTAAACTTCCTGAGTTTGAAATTTGTT 960

DB 901 AGAGTTTGAAGGAATTCAAATATGAAGACATTTCTAAACTTCGCTGAGTTTGAAATTTGTT 960

QY 961 AACATACCTTTAAATCTTAAAGCATCAAAATTTGGTGTTCG 1002
DB 961 AACATACCTTTAAATCTTAAAGCATCAAAATTTGGTGTTCG 1002

RESULT 2
AX85550
ID AX85550 standard; cDNA; 1102 BP.
XX
AC AX85550;
XX
DT 07-SEP-1999 (first entry)
XX
DE cDNA of a gene coding for a mouse deleted kin17 protein.
XX
KW kin17 protein; cell proliferation; fertility; hyperproliferative disease;
KW protein interaction; curved DNA; HIV replication; HIV integration;
KW repair enzyme; ss.
XX
OS Mus sp.
XX
PN FR2772046-A1.
XX
PD 11-JUN-1999.
XX
PF 09-DEC-1997; 97FR-00015536.
XX
PR 09-DEC-1997; 97FR-00015536.
XX
(COMS) COMMISSARIAT ENERGIE ATOMIQUE.
XX
PA Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MUF;
PI WPI; 1999-359999/31.
XX
DR New DNA coding for human kin17 protein - useful for controlling cell
XX proliferation or fertility.
PT
PS Claim 4; Page 31; 69pp; French.
XX
CC The present sequence encodes a mouse kin17 protein with amino acids 129-
CC 228 deleted. The mammalian kin17 protein is useful for preparing a
CC medicament for controlling cell proliferation or for controlling
CC fertility. The medicaments can also be used to treat hyperproliferative
CC diseases. Fragments between amino acids 55 and 235 (preferably between
CC amino acids 129 and 228) of a mammalian kin17 protein are useful for
CC regulating the interaction between proteins and curved DNA. The fragment
CC can be used to block replication of HIV or its integration into the human
CC genome or to target repair enzymes to curved DNA sites. Expression
CC vectors for kin17 can be used for controlling cell proliferation
XX
SQ Sequence 1102 BP; 373 A; 205 C; 272 G; 252 T; 0 U; 0 Other;

Query Match 70.3%; Score 704.6; DB 2; Length 1102;
Best Local Similarity 86.1%; Pred. No. 1.6e-183;
Matches 819; Conservative 0; Mismatches 119; Indels 13; Gaps 3;

QY 58 ATGGGGAAGTCGATTTTCTTACTCCCAAGGCTATCGCCAACAGGATCAAGTCCCAAGGG 117
DB 1 ATGGGCAAGTCGATTTTCTGAGCCCAAGGCCATCGCCATAGAAATTAAGTCCCAAGGG 60

QY 118 CTGCAGAAAGCTACGCTGGTATTTGCCAGATGTGCCAAGACAGTGCCTGGGACGAGATGCG 177
DB 61 CTCCAGAAAGCTTCGCTGGTACTGCCAGATGTGCCAAAAGCAATGCGCGACGAGAATGCG 120

QY 178 TTTAAGTGTCAATGATGTCGGAATCTCATCAGAGCAACTATTCGTGGCTTCAGAAAT 237
DB 121 TTTAAGTGTCAATGATGTCGGAATCTCATCAAGACCAACTGTTGCTGGCTTCAGAAAC 180

QY 238 CCTCAGCAGTATTATCGATTATTTTTCAGAGGAATTCGAAATGACTTTCTAGAACTTCTC 297
DB 238 CCTCAGCAGTATTATCGATTATTTTTCAGAGGAATTCGAAATGACTTTCTAGAACTTCTC 297

```

Db 181 CCTCAGCAGTTTATGATTTATTTTTCAGAGGAATTCGAAATGACTTTCTGGAACCTCTG 240
Qy 298 AGAGAGCGTTTGGCACTAAAGGGTCCACAACAATTGTCTACAAAGNATACATCAGC 357
Db 241 AGCGAGCGTTTGGCACTAAAGGGTCCACAACAATTGTCTACAAATGAATACATCAGC 300
Qy 358 CACCGAGAGCAGATCCACATGAATGCCACATCAGTCAGTGGGMAACTCTGACTGATTTTACTAAG 417
Db 301 CACCGAGAGCAGATCCACATGAACGTACCAGTGGGAGACATGACCGACTTTTACCAAG 360
Qy 418 TGGCTGGGACAGAGAGGCTTTGTGCAAA-----AGTGCATCAGAGACGATAGGAAGTTCA 471
Db 361 TGGCTGGGACAGAGAGGCTTTGTGTAAGGTACCAGTGCATGAGCTGTCTGGGAGCGCA 420
Qy 472 GCATCAGTGAACGAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAGAAAGAA 531
Db 421 GCATCCGGGAAACGGAAGAGTCTTTCACAGAGCTCCGCCAGCCT-----GCGAAGAG 474
Qy 532 AAGAAATCTCAGCTGGATGAATCATCGAGATTGAAGAGGAAAGAAAGAACTGCCCGA 591
Db 475 AAGAAATCTCAGCTGGATGAATCATCGAGATTGAAGAGGAAAGAAAGAAAGAGCGACGG 534
Qy 592 ACAGACTACTGGCTACAGCGCTGAAATTTATTTGAAATTTATAACCAAGAAACTGGGAG 651
Db 535 ACAGCGCTGTTTACAGCGCGGGATCGTTGTGAAATTTATAACGAAGAGCTTGGGGAG 594
Qy 652 AATATCATAGAAAAAGGCTATTTTAAAGAGATTAATTGACAAATATACAGCTGTGTG 711
Db 595 AATATCATCAAGAAAGAGGGTGGTTAAGGAAGTGAATGACAGGTACACAGCTGTGGTA 654
Qy 712 AAGATGATGATCTGAGACAGCTGAACTGAACTTGAACCTGACAGACTCAATTTAGAGACAGTAA 771
Db 655 AAGATGATGATCTGAGACAGCTGAACTGAACTGACAGACTCAATTTAGAGACAGTAA 714
Qy 772 CCAGCACAGGAAAGAAATTTAGTTTAAATGGAGGCTACAGAGGAATGAAGGTACC 831
Db 715 CCGGCCCGGGGAAAGGGTCTAGTTTAAATGGAGGCTACAGAGGAATGAAGGCACT 774
Qy 832 CTAGAATCCATCAATGAGAAGACTTTTTCAGCTACTATCGTCAATGAACTGGCCCTTTA 891
Db 775 CTCGAATCCATCAATGAGAAGGCTTTTTCAGCCACGATAGTCAATGAACTGGACCTTTG 834
Qy 892 AAAGGACGAGAGTTGAAGGAATTCATATGAACACATTTCTAAACTTGGCTGAGTTGA 951
Db 835 AAAGGACGAGAGTTGAAGGAATTCATATGAACACATATTCATATGAACACATATTCAT 894
Qy 952 AAATTTGTTAACTAACCTTTAAATCTTAAAGCATCAAAATTTGGTGTTCG 1002
Db 895 AAATTTGTTAACTAACCTTTAAAGCATCAAAATTTGGTGTTCG 944

```

RESULT 3

AA85549
ID AA85549 standard; cDNA; 1296 BP.

XX AC
XX AC
XX AC

DT 07-SEP-1999 (first entry)

DE cDNA of a gene coding for the human kin17 protein.

KW Human; kin17 protein; cell proliferation; fertility;
KW hyperproliferative disease; protein interaction; curved DNA;
KW HIV replication; HIV integration; repair enzyme; ss.

XX OS Homo sapiens.

XX PN FR272046-A1.

XX PD 11-JUN-1999.

XX PF 09-DEC-1997; 97FR-00015536.

```

PR 09-DEC-1997; 97FR-00015536.
XX (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.
XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
XX WPI; 1999-359999/31.
DR New DNA coding for human kin17 protein - useful for controlling cell
XX proliferation or fertility.
PT Claim 1; Page 30-31; 69pp; French.
XX The present sequence encodes a human kin17 protein. The mammalian kin17
XX protein is useful for preparing a medicament for controlling cell
XX proliferation or for controlling fertility. The medicaments can also be
XX used to treat hyperproliferative diseases. Fragments between amino acids
XX 55 and 235 (preferably between amino acids 129 and 228) of a mammalian
XX kin17 protein are useful for regulating the interaction between proteins
XX and curved DNA. The fragment can be used to block replication of HIV or
XX its integration into the human genome or to target repair enzymes to
XX curved DNA sites. Expression vectors for kin17 can be used for
XX controlling cell proliferation
XX SQ Sequence 1296 BP; 453 A; 240 C; 310 G; 293 T; 0 U; 0 Other;

Query Match 69.5%; Score 696.4; DB 2; Length 1296;
Best Local Similarity 77.2%; Pred. No. 3.2e-181;
Matches 1001; Conservative 0; Mismatches 1; Indels 294; Gaps 1;

Qy 1 TGATTGAGCTCGGTACCGGGGATCCGATTAGAAAGTATCGTCCGCTGCGCCATG 60
Db 1 TGATTGAGCTCGGTACCGGGGATCCGATTAGAAAGTATCGTCCGCTGCGCCATG 60
Qy 61 GGGAGTTCGATTTTCTTACTCCAGGCTATGCCACAGGATCAAGTCCAGGGGCTG 120
Db 61 GGGAGTTCGATTTTCTTACTCCAGGCTATGCCACAGGATCAAGTCCAGGGGCTG 120
Qy 121 CAGAGCTACGCTGGTATTCGCCAGATGTGCAGAGCAGTCCGGGACGAGAAATGGCTTT 180
Db 121 CAGAGCTACGCTGGTATTCGCCAGATGTGCAGAGCAGTCCGGGACGAGAAATGGCTTT 180
Qy 181 AAGTGTCAATTTGTTCGGAATCTCATCAGAGACAACTATTGTGGCTTCAGAAATTCCT 240
Db 181 AAGTGTCAATTTGTTCGGAATCTCATCAGAGACAACTATTGTGGCTTCAGAAATTCCT 240
Qy 241 CAGAGTTTATGGAATTTTTCAGAGGAATTCGGAATTCAGAAATTCAGAAATTCCT 300
Db 241 CAGAGTTTATGGAATTTTTCAGAGGAATTCGGAATTCAGAAATTCAGAAATTCCT 300
Qy 301 AGAGCTTTGCACTAAAGGGTCCACAACAATTGTCTACAAAGATACATCAGGCAC 360
Db 301 AGAGCTTTGCACTAAAGGGTCCACAACAATTGTCTACAAAGATACATCAGGCAC 360
Qy 361 CGAGAGCACATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTTTACTAAGTGG 420
Db 361 CGAGAGCACATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTTTACTAAGTGG 420
Qy 421 CTGGGACAGAGAGGCTTTGTGCAAA----- 444
Db 421 CTGGGACAGAGAGGCTTTGTGCAAAAGTGGACGAGACACCAAAAGCTGGTATATTCAGTAC 480
Qy 445 ----- 444
Db 445 ----- 444
Qy 481 ATAGACAGGACCCAGAAACTATCCCGGGCAACTGGAACTGGAGAAAGAAAGAACGAG 540
Db 445 ----- 444
Qy 541 GACCTTGATGATGAAGAAAGAACTGCCAAATTTATTGAAGAGCAAGTGAGAGAGGCTG 600
Db 445 ----- 444
Qy 601 GAAGGAAGGAACAGGAGGCTCCTACTTTTACGGAATTTAAGCAGAGAAATGATGAAGAG 660
Db

```



```

QY 445 ----- 444
Db 661 AAAGTCACGTTTAATTTGAGTAAGAGCATGTAGTCTATCGGAGCAACATCTTCCAAG 720
QY 445 -----AGTGCACTGAAGCATAGGAAGTTTCAGCATCAGTGAACGA 486
Db 721 TCAGTACTCTGGGACCGAGTGCATCGAAGCATAGGAAGTTTCAGCATCAGTGAACGA 780
QY 487 AAAGAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAAAGAAAAAGAAATCTGCACGTG 546
Db 781 AAAGAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAAAGAAAAAGAAATCTGCACGTG 840
QY 547 GATGAATCATGGAGATTGAAGGAAAAAGAAAAAGAACTGCCGAAACAGACTACTGGCTA 606
Db 841 GATGAATCATGGAGATTGAAGGAAAAAGAAAAAGAACTGCCGAAACAGACTACTGGCTA 900
QY 607 CAGCCTGAAATATTCTGAAAAATTATAACCAAGAACTGGAGAGAAATATCATAGAAA 666
Db 901 CAGCCTGAAATATTCTGAAAAATTATAACCAAGAACTGGAGAGAAATATCATAGAAA 960
QY 667 AAGGCTATTGTTAAGGAAGTAATTGACAAATATACAGCTGTTGTGAAGATGATTGATCT 726
Db 961 AAGGCTATTGTTAAGGAAGTAATTGACAAATATACAGCTGTTGTGAAGATGATTGATCT 1020
QY 727 GGAGACAAGCTGAACTTGACCAAGACTCAATTTAGAGACAGTAATTCAGCACCAGGAAA 786
Db 1021 GGAGACAAGCTGAACTTGACCAAGACTCAATTTAGAGACAGTAATTCAGCACCAGGAAA 1080
QY 787 AGAATTTCTAGTTTTTAATCGAGCTACAGAGGAATGAAGTACCTCAAGATCCATCAAT 846
Db 1081 AGAATTTCTAGTTTTTAATCGAGCTACAGAGGAATGAAGTACCTCAAGATCCATCAAT 1140
QY 847 GAGAAGACTTTTTCAGCTACTATCGTCAATGAAACTGGCCCTTTAAAGAGCGCAGAGTT 906
Db 1141 GAGAAGACTTTTTCAGCTACTATCGTCAATGAAACTGGCCCTTTAAAGAGCGCAGAGTT 1200
QY 907 GAAGGAATTCATATGAAGACATTTCTAACTTGCCTGAGTTTGAATAATTTGTTAAACAAT 966
Db 1201 GAAGGAATTCATATGAAGACATTTCTAACTTGCCTGAGTTTGAATAATTTGTTAAACAAT 1260
QY 967 ACCTTTAAATCTTAAAGCATCAATTTGTTGCTCGC 1002
Db 1261 ACCTTTAAATCTTAAAGCATCAATTTGTTGCTCGC 1296

```

RESULT 4

AA85552
ID AAX85552 standard; cDNA; 1002 BP.

XX AAX85552;

XX 07-SEP-1999 (first entry)

XX Probe derived from cDNA of a gene coding for the human kin17 protein.

XX Human; kin17 protein; cell proliferation; fertility;

XX hyperproliferative disease; protein interaction; curved DNA;

XX HIV replication; HIV integration; repair enzyme; probe; ss.

XX Homo sapiens.

XX FR2772046-A1.

XX 11-JUN-1999.

XX 09-DEC-1997; 97FR-00015536.

XX 09-DEC-1997; 97FR-00015536.

XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;

XX DR WPI; 1999-359999/31.
XX PT New DNA coding for human kin17 protein - useful for controlling cell proliferation or fertility.
XX PS Claim 7; Page 32-33; 69pp; French.
XX CC The present sequence is derived from nucleotides 207-1208 of AAX85549, and is a probe used to isolate human kin17 cDNA. The mammalian kin17 protein is useful for preparing a medicament for controlling cell proliferation or for controlling fertility. The medicaments can also be used to treat hyperproliferative diseases. Fragments between amino acids 55 and 235 (preferably between amino acids 129 and 228) of a mammalian kin17 protein are useful for regulating the interaction between proteins and curved DNA. The fragment can be used to block replication of HIV or its integration into the human genome or to target repair enzymes to curved DNA sites. Expression vectors for kin17 can be used for controlling cell proliferation

XX SQ Sequence 1002 BP; 374 A; 179 C; 235 G; 214 T; 0 U; 0 Other;

Query Match 46.9%; Score 470; DB 2; Length 1002;

Best Local Similarity 100.0%; Pred. No. 7.5e-119; Mismatches 0; Gaps 0;

Matches 470; Conservative 0; Indels 0; Gaps 0;

QY 445 AGTGCACTGAAGACGATAGGAAGTTTCAGCATCAGTGAACCGAAAAAGAAATCTTCCAGAGC 504

Db 533 AGTGCACTGAAGACGATAGGAAGTTTCAGCATCAGTGAACCGAAAAAGAAATCTTCCAGAGC 592

QY 505 TCACTCAGTCTAAAGAAAAAGAAAAAGAAATCTGCATCGATGAATCATGAGATT 564

Db 593 TCACTCAGTCTAAAGAAAAAGAAAAAGAAATCTGCATCGATGAATCATGAGATT 652

QY 565 GAAGAGGAAAAAGAAAGAACTGCCGAAACAGACTACTGGCTACAGCTGAAATTTATG 624

Db 653 GAAGAGGAAAAAGAAAGAACTGCCGAAACAGACTACTGGCTACAGCTGAAATTTATG 712

QY 625 AAAATTATAACCAAGAACTGGGAGAGAAATATATAAGAAAAAGGCTATTGTTAAGGAA 684

Db 713 AAAATTATAACCAAGAACTGGGAGAGAAATATATAAGAAAAAGGCTATTGTTAAGGAA 772

QY 685 GTAATTGACAAATATACAGCTGTTGTGAAGATGATTGATCTGGAGACAGCTGAACTT 744

Db 773 GTAATTGACAAATATACAGCTGTTGTGAAGATGATTGATCTGGAGACAGCTGAACTT 832

QY 745 GACCAGACTCTATTTAGACACAGTAATTCAGCACCAGGAAAAAGAAATCTAGTTTTTAAT 804

Db 833 GACCAGACTCTATTTAGACACAGTAATTCAGCACCAGGAAAAAGAAATCTAGTTTTTAAT 892

QY 805 GGAGGCTACAGAGGAAATGAAGGTACCTTAGAATCCATCAATGAGAAGACTTTTTTCAGCT 864

Db 893 GGAGGCTACAGAGGAAATGAAGGTACCTTAGAATCCATCAATGAGAAGACTTTTTTCAGCT 952

QY 865 ACTATCGTCATTGAACTGGCCCTTTAAAGAGCGCAGAGTTGAAGGAAT 914

Db 953 ACTATCGTCATTGAACTGGCCCTTTAAAGAGCGCAGAGTTGAAGGAAT 1002

RESULT 5

ADQ56943

ID ADQ56943 standard; DNA; 679 BP.

XX ADQ56943;

XX 21-OCT-2004 (first entry)

XX Novel canine microarray-related DNA sequence SeqID8245.

XX canine microarray; drug screening; toxicity assay;

XX environmental pollutant; cellular response; gene expression profile;

XX toxic response; liver necrosis; fatty liver disease;

XX protein adduct formation; hepatitis; dog; dr.

Best Local Similarity 66.1%; Pred. No. 4.7e-104;		Matches 819; Conservative 0; Mismatches 119; Indels 301; Gaps 3;	
QY	58	ATGGGAAGTCGGATTTCTTACTCCCAAGGCTATCGCCAAAGGATCAAGTCCAAAGGG	117
Db	1	ATGGCAAGTCGGATTTCTGAGCCCAAGGCCATCGCCAATAGAATTAACTCCAAGGG	60
QY	118	CTGCAGAACCTAGCTGGTATTCGCCAGATGTGCCAAGAGAGTGCCTGGGACGAGATGGC	177
Db	61	CTCCAGAAGCTTGGCTGGTACTGCCAGATGTGCCAAAAGCAATGCCCGACGAGATGGC	120
QY	178	TTTAACTGTCTATGTATGTCGGAATCTCATCAGACACAATATTGTGCTTCAGAAAT	237
Db	121	TTTAACTGTCTATGTATGTCGGAATCTCATCAAGACAACTGTTGCTGCTCAGAAAC	180
QY	238	CCTCAGCAGTTATGGATTAATTTTTCAGAGGAATTCGAAATGACTTTTCTAGAATTTCTC	297
Db	181	CCTCAGCAGTTATGGATTAATTTTTCAGAGGAATTCGAAATGACTTTTCTGGAATTTCTG	240
QY	298	AGGAGAGCGTTTGGCACTAAAGGGTCCACAACAACATTTGTCTAACGAATACATCAGC	357
Db	241	AGGCGAGCGTTTGGCACTAAAGGGTCCACAACAACATTTGTCTAACGAATACATCAGC	300
QY	358	CACCGAGAGCACATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTTTACTAAG	417
Db	301	CACCGAGAGCACATCCACATGAACGCTACCAGTGGGAGACACTGACCGACTTTTACCAAG	360
QY	418	TGGCTGGGCAGAGAAGCGTTGTGCAA-----	443
Db	361	TGGCTGGGCAGAGAGGCTTGTGTAAAGTGGATGAGACACCGAAGGCTGGTACATTCAG	420
QY	444	-----	443
Db	421	TACATAGACAGAGACCCAGAAACCATCCGTGGCAACTGGAAATTAGAAAAAAGAGAAG	480
QY	444	-----	443
Db	481	CAAGATCTGGACGATGAAGAAAAAACTGCCAAGTTCATTGAGGAGCAGGTGAGAGAGGC	540
QY	444	-----	443
Db	541	CTGGAAGGGAAGAGCAGGAGAGACACCTGTTTTTACAGAACTTAGCCGAGAAATGAGGAA	600
QY	444	-----	443
Db	601	GAAAAATTACGTTCAATCTGAATAAAGGAGCGGTGGCTCAGCGGGAGCTACAACATCC	660
QY	444	-----	483
Db	661	AAGTCAAGCTCTTTGGGACCAAGTGCACCTGAAGCTCTGGGGAGCGCAGCATCCGGGAA	720
QY	484	CGAAAAAATCTTCCAGAGCTCAACTCAGTCTTAAAGAAAAAGAAAAAGAAATCTGCA	543
Db	721	CGAAAAAGCTCTTACAGAGCTCCGCCAGCT-----GCGAAGAAAGAAAGTCCGCC	774
QY	544	CTGGATGAATCATGTAGATTTGAAGAGGAAAAAGAAAAAGAACTCCCGCAACAGACTACTGG	603
Db	775	CTGGATGAGATCATGTAGCTCGAAGAGGAAAAAGAAAAAGGACCGCAGCAGACGCTCG	834
QY	604	CTACAGCCTGAATTAATTTGTAAAAATTATAACCAAGAACTGGGAGAGAAATATCATAAG	663
Db	835	TTACAGCCGGGATCGTGTGTAAAAATTATAACGAGAGCTTTGGGAGAAATATCACAAG	894
QY	664	AAAAAGGCTATTGTTAAGGAAGTAATTGACAAATATACAGCTGTTGTGAAGATGATTGAT	723
Db	895	AGAAAGGGTCTGTTAAGGAGTGAATTGACAGGTACACAGCTGTGTTAAGATGACTGAC	954
QY	724	TCTGGAGACAGCTGAAAATTGACAGCTCATTTAGACACAGTAATTTCCAGCACCAGGA	783
Db	955	TCTGGAGACAGGCTGAAAATTGACAGCTCATTTAGACACAGTCAATTCGGGCCCGGG	1014
QY	784	AAAAAGAAATCTAGTTTTTAATGAGGCTACAGAGGAATGAAGTACCTTAGAATCCATC	843

Db	1015	AAAAGGGTCTAGTTTTTAAATGAGGCTACAGAGAAATGAAGGCACCTCTCGAATCCATC	1074
QY	844	AATGAGAGACTTTTTCAGCTACTATCGTCAATGAACTGGCCCTTTAAAAGGACGCAGA	903
Db	1075	AATGAGAAAGGCTTTTTCAGCCACGATAGTCAATGAACTGGACCTTTGAAAGGACGCAGA	1134
QY	904	GTTGAAGGAATTCATATGAAGACATTTCTAAACTTGCCTGAGTTTGAAAAATTTGTTAAC	963
Db	1135	GTTGAAGGATTCATATGAAGACATATCTAACTTGTGTGAGTTTGAAAAATTTGATAAC	1194
QY	964	AATACCTTTAAATCTTAAAGCATCAAAATTTGTTTCG 1002	
Db	1195	AACACATTGAACT-GTGAAGCATCAAAATTTGTTTACG 1232	
RESULT 7			
AAQ79936			
ID	AAQ79936	standard; cDNA; 1458 BP.	
XX	AAQ79936;		
XX	25-MAR-2003 (revised)		
DT	06-SEP-1995 (first entry)		
XX	Murine Kin17 cDNA.		
DE			
XX	chromosomal rearrangement; kin17 protein; SOS DNA repair system; RecA;		
KW	genotoxic agent; zinc finger; DNA binding protein; ss.		
OS	Mus musculus.		
XX			
FT	Key	Location/Qualifiers	
FT	primer_bind	complement(1..21)	
FT		/*tag= b	
FT	misc_feature	/note= "Oligo L (AAQ79946) binds to complementary strand"	
FT		22..1434	
FT		/*tag= l	
FT		/label= kin17_cDNA	
FT		/note= "nucleotides 1-1414; the genomic DNA contains at least two introns within this sequence, see Comments"	
FT	primer_bind	32..49	
FT		/*tag= c	
FT		/note= "Oligo C (AAQ79938) binding site"	
FT	CDS	46..1221	
FT		/*tag= a	
FT		/product= "Kin17"	
FT	primer_bind	/note= "N's in the sequence denote illegible residues"	
FT		complement(67..86)	
FT		/*tag= d	
FT	primer_bind	/note= "Oligo S (AAQ79947) binds to complementary strand"	
FT		274..297	
FT		/*tag= e	
FT	primer_bind	/note= "Oligo D (AAQ79939) binding site"	
FT		complement(339..360)	
FT		/*tag= f	
FT	primer_bind	/note= "Oligo K (AAQ79945) binds to complementary strand"	
FT		451..474	
FT		/*tag= g	
FT	primer_bind	/note= "Oligo J (AAQ79944) binding site"	
FT		complement(550..567)	
FT		/*tag= h	
FT	primer_bind	/note= "Oligo E (AAQ79940) binds to complementary strand"	
FT		802..825	
FT		/*tag= i	
FT	primer_bind	/note= "Oligo F (AAQ79941) binding site"	
FT		complement(839..862)	
FT		/*tag= j	
FT	primer_bind	/note= "Oligo G (AAQ79942) binds to complementary strand"	
FT		complement(1435..1458)	
FT		/*tag= k	
FT	primer_bind	/note= "Oligo B (AAQ79937) binds to complementary strand"	
XX	FR2706487-Al.		
PN			

XX 29-SEP-1999; 99US-0157137P.
PR 03-NOV-1999; 99US-0163280P.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
XX WPI; 2001-235357/24.
DR P-PSDB; AAG74690.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers.
XX
PS Claim 1; Page 2977-2978; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene therapy
CC and vaccine production. N and P may be used in the prevention, diagnosis
CC and treatment of diseases associated with inappropriate P expression. For
CC example, N and P may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of P by expressing inactive proteins or to
CC supplement the patients own production of P. Additionally, N may be used
CC to produce the colon cancer-associated Ps, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the proteins. N and P
CC can be used in the prevention, diagnosis and treatment of colorectal
CC carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent
CC sequences used in the exemplification of the present invention. N.B.
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC time of publication, meaning no sequences are present for SEQ ID NO:1027
CC to 1052, 7921 and 7922
XX
XX Sequence 591 BP; 199 A; 95 C; 114 G; 182 T; 0 U; 1 Other;
SQ
Query Match 26.7%; Score 267.2; DB 4; Length 591;
Best Local Similarity 95.3%; Pred. No. 4.9e-63;
Matches 286; Conservative 0; Mismatches 13; Indels 1; Gaps 1;
QY 703 GCTGTTGTAAGATGATGATCTGAGACAAAGCTGAAACTTGACAGACTCATTTAGAG 762
DB 28 GCTGCAGGAATTCGGCAGCAGCTGGAGACAAGCTGAAACTTGACAGACTCATTTAGAG 87
QY 763 ACAGTAATTCAGCACCAGGAAAAAGAAATCTAGTTTTTAAATGGAGGCTACAGAGGAAT 822
DB 88 ACAGTAATTCAGCACCAGGAAAAAGAAATCTAGTTTTTAAATGGAGGCTACAGAGGAAT 147
QY 823 GAAGGTACCTAGATCCATCAATGAGAAGAATTTTTCAGCTACTATCGTCATTGAAACT 882
DB 148 GAAGGTACCTAGATCCATCAATGAGAAGAATTTTTCAGCTACTATCGTCATTGAAACT 207
QY 883 GGCCTTTTAAAGGACGAGTTGAAGGAATTCATATGAGACATTTCTAAACTTGGC 942
DB 208 GGCCTTTTAAAGGACGAGTTGAAGGAATTCATATGAGACATTTCTAAACTTGGC 267
QY 943 TGAGTTTGAATAATTTGTTAAACATACCTTTTAAATCTTAAAGCATCAAAATTTGGTTTCG 1002
DB 268 TGAGTTTGAATAATTTGTTAAACATA-CATTAAATCTTAAAGCATCAAAATTTGGTTTCG 326
RESULT 9
ADP22451
ID ADP22451 standard; DNA; 1372 BP.
XX
XX AC ADP22451;
XX
XX 12-AUG-2004 (first entry)
XX
XX DE Sea-squirt (Ciona intestinalis) zinc finger protein coding sequence #13.
XX
XX KW sea-squirt; zinc finger protein; gene detection; drug development;

zinc finger protein-associated disease; gene; ds.
KW
XX Ciona intestinalis.
OS
XX JP2004057126-A.
PN
XX 26-FEB-2004.
PD
XX 31-JUL-2002; 2002JP-00222484.
PF
XX 31-JUL-2002; 2002JP-00222484.
PR
XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
PA
XX WPI; 2004-208711/20.
DR P-PSDB; ADP22452.
DR
XX Novel gene encoding zinc finger protein, useful as probe in gene
PT detecting instruments and in development of drug for treating zinc finger
PT protein associated diseases.
XX
XX Claim 3; SEQ ID NO 25; 972pp; Japanese.
PS
XX The invention comprises the amino acid and coding sequences of sea-squirt
CC (Ciona intestinalis) zinc finger proteins. The DNA and protein sequences
CC of the invention are useful in a gene detecting instrument. The DNA and
CC protein sequences of the invention are useful in the development of drugs
CC for the treatment of zinc finger protein-associated diseases. The present
CC DNA sequence encodes a sea-squirt zinc finger protein of the invention.
XX
XX Sequence 1372 BP; 500 A; 224 C; 286 G; 362 T; 0 U; 0 Other;
SQ
Query Match 20.1%; Score 201.4; DB 12; Length 1372;
Best Local Similarity 63.1%; Pred. No. 1e-44;
Matches 310; Conservative 0; Mismatches 181; Indels 0; Gaps 0;
QY 61 GGGAGTCGGATTTTCTTACTCCCAAGGCTATCGCAACAGATCAAGTCCCAAGGGCTG 120
DB 52 GAGAAACAGGATTTCTCACTGCCAAGGCTATCGGCAACCGTATCAATCAAGGACTT 111
QY 121 CAGAAAGCTACGCTGGTATTTGCCAGATGTGCCAGAACAGTCCGGGACGAGATGGCTTT 180
DB 112 CAAAAGTTGCGATGCTATTTGTCAAATGTGCCAAAACAGTGCAGAGATGAGAACGGTTT 171
QY 181 AAGTCTATTGATGTCGGAATCTCATCAGAGAACATATTGCTGCTGCTTCAGAAAATCCT 240
DB 172 AAATGCCATTGATGTCGGAATCATCAGGCGCAATTTGTTAAATGCTGGTGAATAATCCT 231
QY 241 CAGCAGTTTATGGATTTATTTTTCAGAGGAATTCGGAATGACTTTCTAGAACTTCTCAGG 300
DB 232 GGAGCTTTTTCAGGTTTCTTTTCAAGTTCCTTTTCAAGACTTTTATGCACTTTTGAA 291
QY 301 AGACCTTTTGGCCTTAAAGGGTCCACAAACAACTTGTCTACAAAGAAATACATCAGCCAC 360
DB 292 ACACCTTTTGGGACTTAAAGAGTTTCATAACAATATTGTATACAGATATATTTCCTCAT 351
QY 361 CGAGAGCATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTTTACTAAGTGG 420
DB 352 AAAGAACACATTCATATGAACGCTACAAAATGGGTGAGCTTAACTGCTTCACAAAATGG 411
QY 421 CTGGGACAGAGAGGCTTCTGCAAAAGTGCATCGAAGACGATAGGAAGTTTCAGCATCAGTG 480
DB 412 TTGGGTAGAGAGGGACACTGTAAAGTGGATCAACCGAAAGGGTGTGTTTCATACATAC 471
QY 481 AAACGAAAAGAAATCTTCCAGAGCTCAACTCAGTCTTAAAGAAAAGAAAAGAAATCT 540
DB 472 ATTGATAGAGACCCATTGCTGCTGCCACAGACAGAAAGAACTTGTATTAACAAAGCAAG 531
QY 541 GCACCTGGATGA 551
DB 532 GAGAGAGATGA 542

RESULT 10
ADT17629
ID ADT17629 standard; cDNA; 1713 BP.
XX
AC ADT17629;
XX
DT 13-JAN-2005 (first entry)
XX
DE Plant cDNA, Seq ID 2955.
XX
KW Plant; ss; gene; transgenic; cold tolerance; growth rate;
KW drought tolerance; disease resistance; galactomannan production;
KW plant growth regulator; heat tolerance; herbicide tolerance;
KW lignin production; extreme osmotic condition tolerance;
KW pathogens resistance; pest resistance; yield improvement; seed oil yield;
KW seed protein yield.
XX
OS Viridiplantae.
XX
PN US2004216190-A1.
XX
PD 28-OCT-2004.
XX
PF 18-DEC-2003; 2003US-00739930.
XX
PR 28-APR-2003; 2003US-00424599.
PR 28-APR-2003; 2003US-00425115.
XX
PA (KOVA/) KOVALIC D K.
XX
PI Kovalic DK;
XX
DR WPI; 2004-757369/74.
XX
PT New recombinant DNA constructs useful in the field of biochemistry and
PT genetics, and in particular for producing transgenic plants with improved
PT biological characteristics.
XX
PS Claim 1; SEQ ID NO 2955; 14pp; English.
XX
CC The invention relates a recombinant DNA construct comprising a
CC polynucleotide having any of 5544 nucleotide sequences (cDNAs SEQ ID NO:
CC 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences
CC (SEQ ID NO: 5545-11088). The cDNAs and proteins are from corn, soybean,
CC Arabidopsis, wheat and rape but the specification does not indicate which
CC sequences is derived from which organism. Also included is a method of
CC producing a plant having an improved property, comprising transforming a
CC plant with a recombinant DNA construct comprising a promoter region
CC functional in a plant cell operably joined to a polynucleotide encoding a
CC polypeptide associated with the property, and growing the transformed
CC plant. The property is selected from improving plant cold tolerance, for
CC manipulating growth rate in plant cells by modification of the cell cycle
CC pathway, for improving plant drought tolerance, for providing increased
CC resistance to plant disease, for galactomannan production, for production
CC of plant growth regulators, for improving plant heat tolerance, for
CC improving plant tolerance to herbicides, for increasing the rate of
CC homologous recombination in plants, for lignin production, for improving
CC plant tolerance to extreme osmotic conditions, for improving plant
CC tolerance to pathogens or pests, for yield improvement by modification of
CC photosynthesis, for modifying seed oil yield and/or content, for
CC modifying seed protein yield and/or content, for yield improvement by
CC modification of carbohydrate, nitrogen or phosphorus use and/or uptake
CC and for yield improvement by providing improved plant growth and
CC development under at least one stress condition. The polynucleotide may
CC also encode a plant transcription factor. The methods and compositions of
CC the present invention are useful in the field of biochemistry and
CC genetics, in particular for producing transgenic plants with improved
CC biological characteristics such as increased yield, improved nitrogen
CC flow, increasing plant tolerance to cold or heat, improving plant
CC tolerance to extreme osmotic and drought conditions, and improving plant
CC tolerance to plant pests or pathogens. They can also be used in physical
CC arrays of molecules, plant breeding markers, computer-based storage and
CC analysis systems. The present sequence is one of the 5544 plant cDNA

CC sequences of the invention. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20040216190.

XX
SQ Sequence 1713 BP; 412 A; 406 C; 522 G; 373 T; 0 U; 0 Other;
Query Match 17.6%; Score 176; DB 13; Length 1713;
Best Local Similarity 63.9%; Pred. No. 1.1e-37;
Matches 266; Conservative 0; Mismatches 150; Indels 0; Gaps 0;
QY 54 CGCCATGGGGAAGTCGGATTCTTCTTACTCCCAAGGCTATCGCAACAGAGATCAAGTCCAA 113
DB 171 CGCGATGGGGAAGACGACGAGTTCCTGACGCGAAGCGGATCGCAACCCGATCAAGCGGAA 230
QY 114 GGGGCTGCAGAGACTACGCTGATTTGCCAGATGTGCCAGAGACGAGTGCCTGGGACGAGAA 173
DB 231 GGGGCTGCAGAGAGCTGCGGTGATGTCAGAGATGTGTCAGAGACGAGTGCCTGGGACGAGAA 290
QY 174 TGGCTTTAAGTGTCTATTGTATGTCGGAATCTCATCAGAGACAACTATTGCTGGCTTCAGA 233
DB 291 CGGGTTCAAGTGCCACTGCTATGTCGAGTCCGACCGGCGAGATGACAGGTGTTCCGCAT 350
QY 234 AAATCCTCAGCAGTTTATGGATTATTTTCAGAGGAATTCGGAATGACTTCTAGAACT 293
DB 351 GCGCGCCGACCGCGTCTGTCGAGGGCTTCTCCGAGGAATTCCTCGAGTCTCTCTCCCT 410
QY 294 TCTCAGGAGACGCTTTGGCACTAAAAGGCTGCACACACATTTGCTACAAACGAATACAT 353
DB 411 CATCGCCCGCGCGACCGCCACTCCCGCGTCCGCGCCACCGTGTCTACAAACGAGTACAT 470
QY 354 CAGCCACCGAGAGACATCCACATGAATGCCACTCAGTGGGAACTCTGACTGATTTTAC 413
DB 471 CGCGGACCGGCGACCGCTCCATGAATCTACGCGGTGGCGCCACGCTCACCAGTTCGT 530
QY 414 TAAGTGGCTGGGCGAGAGAGGCTTGTGCAAAAGTGCCTGAGACGATAGGAAGTT 469
DB 531 CAAAGCTCTGGGCGCGAAGGGTACTGCAAGGTTGAGAGACACGCCCAAGGGTGGT 586
RESULT 11
AAA52591
ID AAA52591 standard; cDNA; 1780 BP.
XX
AC AAA52591;
XX
DT 27-SEP-2000 (first entry)
XX
DE cDNA encoding maize KIN17 orthologue, ZmKINH-3.
XX
KW ZmKINH-3; KIN17 orthologue; maize; zinc finger protein; RecA homologue;
KW nuclear localisation; nonhomologous recombination;
KW illegitimate recombination; double stranded DNA binding; curved DNA;
KW homologous gene targeting; transgenic plant; ss.
XX
OS Zea mays.
XX
FH Key Location/Qualifiers
FT CDS 172..1443
FT /tag= a
FT /product= "Maize KIN17 orthologue, ZmKINH-3"
FT /function= "Binds double-stranded curved DNA"
XX
PN WO200024900-A1.
XX
PD 04-MAY-2000.
XX
PF 06-OCT-1999; 99WO-US023280.
XX
PR 27-OCT-1998; 98US-0105802P.
XX
PA (PTON-) PIONEER HI-BRED INT INC.
XX


```
PI Mahajan PB;
XX WPI; 2000-350741/30.
XX Nucleic acids encoding maize KIN17 orthologue proteins useful for
PT preventing illegitimate recombination in cells.
XX Claim 1; Page 77-79; 84pp; English.
XX This sequence represents cDNA encoding the maize KIN17 orthologue ZmKINH-
CC 3. The invention relates to maize KIN17 orthologues (AAB03063- AAB03065),
CC nucleic acids encoding them (AAAS2589-A52591), and expression vectors,
CC transgenic plants and plant seeds comprising nucleotides encoding maize
CC KIN17 orthologues. KIN17 has, until now, been found only in animal
CC (avian, rodent and human) cells, this invention being the first report
CC describing the presence of KIN17 in plants. Murine KIN17 was found to
CC have significant homology to Escherichia coli RecA protein, and contains
CC a zinc finger motif and a nuclear localisation signal. KIN17 binds double
CC -stranded DNA, preferentially binding to curved DNA, and forms
CC intranuclear foci on overexpression in mammalian cells. It is also
CC induced on exposure to gamma or ultraviolet radiation. These findings
CC indicate that KIN17 plays a role in non-homologous (illegitimate)
CC recombination, which occurs at higher rates among higher eukaryotes,
CC particularly plants. Illegitimate recombination in plants is a major
CC impediment to the generation of transgenic crops such as maize. Maize
CC KIN17 orthologue nucleic acid sequences may be used to generate
CC transgenic plants. The transgenic plants generated can be monocots or
CC dicots and are particularly maize, soybean, sunflower, sorghum, canola,
CC wheat, alfalfa, cotton, rice, barley and millet. In particular, antisense
CC sequences may be used to reduce KIN17 levels in embryogenic callus or
CC embryo cells, thereby reducing the amount of non-homologous recombination
CC and enhancing homologous gene targeting
XX SQ Sequence 1780 BP; 467 A; 408 C; 525 G; 380 T; 0 U; 0 Other;
Query Match 17.6%; Score 176; DB 3; Length 1780;
Best Local Similarity 63.9%; Pred. No. 1.2e-37;
Matches 266; Conservative 0; Mismatches 150; Indels 0; Gaps 0;
QY 54 CGCATGGGGAGTCGGATTTCTTACTCCCAAGGCTATCGCAACAGCATCAAGTCCAA 113
DB 168 CGCGATGGGGAGTCGGATTTCTTACTCCCAAGGCTATCGCAACAGCATCAAGTCCAA 227
QY 114 GGGGCTGCAGAGCTAGCTGGTATTTGCCAGATGTCGCAAGCAGTCGCGGACGAGAA 173
DB 228 GGGGCTGCAGAGCTAGCTGGTATTTGCCAGATGTCGCAAGCAGTCGCGGACGAGAA 287
QY 174 TGGCTTTAAGTGTCATTGTATGTCGCAATCTCATCAGAGCACTATTGCTGCTTCAGA 233
DB 288 CGGGTTCAAGTGCACCTGCATGTCGAGTGCACACGCGCAGATCGAGTGTTCGGCAT 347
QY 234 AAATCCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTCGGAATGACTTTCTAGAAT 293
DB 348 GGGCCCGACCGGTCGTCGAGGGCTTCTCCGAGGAATTCCTCGAGTCTCTCTCCCT 407
QY 294 TCTCAGAGAGCGTTTGGCACTAAAGGGTCCACACACATTTGCTTACACGAATACAT 353
DB 408 CATCCGCGCGGACCGCACCTCCCGCGTTCGCGCACCGTGTGCTACACGAGTACAT 467
QY 354 CAGCCACCGAGACACATCCACATGAATGCACCTCAGTGGGAACCTCTGACTGATTTTAC 413
DB 468 CGCGGACCGGACACATCCACATGAATCTACCGGGTGGCCACCGCTCACCGAGTTCGT 527
QY 414 TAAAGTGGCTGGGAGAGAGGGCTTGTGCAAAAGTGCACCTGAAGACGATAGGAAT 469
DB 528 CAAAGTCTCTGGGCGCGAGAGGGTACTGCAAGGTTGAGGACACGCCCAAGGGTGGT 583
RESULT 12
ID AAAS2590
XX AAAS2590 standard; cDNA; 1780 BP.
XX AC AAAS2590;
```

```
XX 27-SEP-2000 (first entry)
XX cDNA encoding maize KIN17 orthologue, ZmKINH-2.
XX ZmKINH-2; KIN17 orthologue; maize; zinc finger protein; RecA homologue;
KW nuclear localisation; nonhomologous recombination;
KW illegitimate recombination; double stranded DNA binding; curved DNA;
KW homologous gene targeting; transgenic plant; ss.
XX Zea mays.
XX Key Location/Qualifiers
CDS 172..1443
FT /*tag= a
FT /product= "Maize KIN17 orthologue, ZmKINH-2"
FT /function= "Binds double-stranded curved DNA"
XX WO200024900-A1.
XX 04-MAY-2000.
XX 06-OCT-1999; 99WO-US023280.
XX 27-OCT-1998; 98US-0105802P.
XX (PION-) PIONEER HI-BRED INT INC.
XX Mahajan PB;
XX WPI: 2000-350741/30.
XX P-PSDB; AAB03064.
XX Nucleic acids encoding maize KIN17 orthologue proteins useful for
PT preventing illegitimate recombination in cells.
XX Claim 1; Page 74-76; 84pp; English.
XX This sequence represents cDNA encoding the maize KIN17 orthologue ZmKINH-
CC 2. The invention relates to maize KIN17 orthologues (AAB03063- AAB03065),
CC nucleic acids encoding them (AAAS2589-A52591), and expression vectors,
CC transgenic plants and plant seeds comprising nucleotides encoding maize
CC KIN17 orthologues. KIN17 has, until now, been found only in animal
CC (avian, rodent and human) cells, this invention being the first report
CC describing the presence of KIN17 in plants. Murine KIN17 was found to
CC have significant homology to Escherichia coli RecA protein, and contains
CC a zinc finger motif and a nuclear localisation signal. KIN17 binds double
CC -stranded DNA, preferentially binding to curved DNA, and forms
CC intranuclear foci on overexpression in mammalian cells. It is also
CC induced on exposure to gamma or ultraviolet radiation. These findings
CC indicate that KIN17 plays a role in non-homologous (illegitimate)
CC recombination, which occurs at higher rates among higher eukaryotes,
CC particularly plants. Illegitimate recombination in plants is a major
CC impediment to the generation of transgenic crops such as maize. Maize
CC KIN17 orthologue nucleic acid sequences may be used to generate
CC transgenic plants. The transgenic plants generated can be monocots or
CC dicots and are particularly maize, soybean, sunflower, sorghum, canola,
CC wheat, alfalfa, cotton, rice, barley and millet. In particular, antisense
CC sequences may be used to reduce KIN17 levels in embryogenic callus or
CC embryo cells, thereby reducing the amount of non-homologous recombination
CC and enhancing homologous gene targeting
XX SQ Sequence 1780 BP; 467 A; 408 C; 525 G; 380 T; 0 U; 0 Other;
Query Match 17.6%; Score 176; DB 3; Length 1780;
Best Local Similarity 63.9%; Pred. No. 1.2e-37;
Matches 266; Conservative 0; Mismatches 150; Indels 0; Gaps 0;
QY 54 CGCATGGGGAGTCGGATTTCTTACTCCCAAGGCTATCGCAACAGCATCAAGTCCAA 113
DB 168 CGCGATGGGGAGTCGGATTTCTTACTCCCAAGGCTATCGCAACAGCATCAAGTCCAA 227
QY 114 GGGGCTGCAGAGCTAGCTGGTATTTGCCAGATGTCGCAAGCAGTCGCGGACGAGAA 173
```


Db 228 GGGGCTGCAGAGCTGGGTGGTACTGCCAGATGTGTCAAGACAGTGGCGGACAGAA 287
 Qy 174 TGGCTTTAAGTGTCAATGTATGTCGGAATCTCATCAGAGACAATATGTGCTTCAGA 233
 Db 288 CGGGTTCAAGTGCACCTGCATGTCGAGTGGCAGTCCGACCGGACAGTGCAGGTTCGGCAT 347
 Qy 234 AATCTCTCAGCAGTTATGATATTTTTCAGAGGAATTCGGAATGACTTTCTAGAACT 293
 Db 348 GGGCCCGGACCGGTGCTCGAGGGCTTCTCGAGGAATTCCTCGAGTCTTCTCTCCCT 407
 Qy 294 TCTCAGGAGACGCTTTTGGCACTAAAGGGTCCACAAATTTGTCTTACACGAATACAT 353
 Db 408 CATCCGCGCGGACCGCCACTCCCGGTGCGCGCCACCGTTGTCTACACGAGTACAT 467
 Qy 354 CAGCCACCGAGACATCCACATGATGCACTCAGTGGGAATCTGTGATGATTTAC 413
 Db 468 CGCGACCGGACACACGTCACATGAACTCTACGCGTGGCGCCACGCTCACCGAGTTGCT 527
 Qy 414 TAAGTGGCTGGCAGACAGGCTTGTGCAAAAGTGCACCTCAAGACGATAGAGTT 469
 Db 528 CAAAGCTCTCGGCGCGAGAGGTACTGCAAGGTTGAGGACACGCCCAAGGGGTGGT 583

RESULT 13

AAA01559
 ID AAA01559 standard; cDNA; 300 BP.

AC AAA01559;

DT 19-MAY-2000 (first entry)

XX Human colon cancer cell line polynucleotide sequence SEQ ID NO:1550.

XX Human; colon cancer; tumour; diagnosis; gene expression product; probe;
 XX detection; cancerous state; metastasis; identification; breast cancer;
 XX oestrogen receptor-positive breast cancer; therapy;
 XX oestrogen receptor-negative breast cancer; lung cancer; ss.

OS Homo sapiens.

XX WO9558675-A2.

XX 18-NOV-1999.

XX 13-MAY-1999; 99WO-US010602.

XX 14-MAY-1998; 98US-0085426P.

XX 15-MAY-1998; 98US-0085537P.

XX 21-OCT-1998; 98US-0105234P.

XX 27-OCT-1998; 98US-0105877P.

XX (CHIR) CHIRON CORP.

XX (HYSE-) HYSEQ INC.

XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;

XX Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;

XX Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;

XX Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;

XX WPI; 2000-126369/11.

XX Polynucleotide library used to determine cancerous states of mammalian

XX cells.

XX Claim 1; Page 553; 1097pp; English.

XX AAA00010 to AAA02716 represent polynucleotides isolated from cDNA

XX libraries constructed from human colon cancer cell lines. The present

CC sample derived from a cell suspected of being cancerous, where detection
 CC of the differentially expressed gene product is correlated with a
 CC cancerous state of the cell from which the test sample was derived. The
 CC polynucleotide sequences can be used in a method for detecting
 CC differentially expressed genes correlated with a cancerous state of a
 CC mammalian cell. The polynucleotides can also be used as probes for
 CC detecting and mapping related genes. They can be used in diagnosis and
 CC prognosis of diseases and disorders (e.g. identification of pre-
 CC metastatic or metastatic cancerous states, stages of cancer, or
 CC responsiveness of cancer to therapy). This is particularly for breast
 CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
 CC negative breast cancer, lung cancer, and colon cancer.

XX SQ Sequence 300 BP; 114 A; 55 C; 71 G; 60 T; 0 U; 0 Other;

Query Match 17.1%; Score 171; DB 3; Length 300;

Best Local Similarity 100.0%; Pred. No. 1.2e-36;

Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 445 AGTGCCTCAAGACGATAGAACTTCAGCATCAGTGAAGAAAAGAAATCTTCCAGAGC 504

Db 130 AGTGCCTCAAGACGATAGAACTTCAGCATCAGTGAAGAAAAGAAATCTTCCAGAGC 189

Qy 505 TCAACTCAGTCTAAAGAAAAGAAAGAAATCTGCATCTGATGAATCATGGAGATT 564

Db 190 TCAACTCAGTCTAAAGAAAAGAAAGAAATCTGCATCTGATGAATCATGGAGATT 249

Qy 565 GAAGAGAAAAGAAAAGAACTGCCGAAACAGACTACTGGCTACAGCCTGAA 615

Db 250 GAAGAGAAAAGAAAAGAACTGCCGAAACAGACTACTGGCTACAGCCTGAA 300

RESULT 14

AD084255
 ID AD084255 standard; cDNA; 1614 BP.

AC AD084255;

DT 21-APR-2005 (first entry)

XX Plant full length insert polynucleotide seqid 2975.

XX plant protectant; plant growth regulant; gene therapy; plant;
 XX recombinant DNA construct; physical array; plant breeding marker;
 XX cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
 XX extreme osmotic condition; pathogen tolerance; pest tolerance;
 XX growth rate; cell cycle pathway; disease resistance;
 XX galactomannan production; lignin production; plant growth regulator;
 XX yield; plant growth; plant development; seed oil; protein yield;
 XX protein content; gene; ss.

OS Unidentified.

XX US2004034888-A1.

XX 19-FEB-2004.

XX 28-APR-2003; 2003US-00425114.

XX 06-MAY-1999; 99US-00304517.

XX 05-NOV-2001; 2001US-00985678.

XX (LIUJ/) LIU J.

XX (ZHOU/) ZHOU Y.

XX (KOVA/) KOVALIC D K.

XX (SCRE/) SCREEN S E.

XX (TABA/) TABASKA J E.

XX (CAOY/) CAO Y.

XX Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;

XX WPI; 2004-180133/17.

PT New recombinant DNA construct, useful for improving plant tolerance to
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT pests, for conferring increased resistance to plant disease, or for
PT improving yield.

XX Claim 1; SEQ ID NO 2975; 15pp; English.

XX The invention describes a recombinant DNA construct comprising a
CC polynucleotide consisting of a sequence encoding an amino acid sequence
CC available in electronic form from the US patent office at
CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC increased resistance to plant disease, for producing galactomannan,
CC lignin or plant growth regulators, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This sequence represents a plant full length insert
CC polynucleotide that can be used in the recombinant DNA construct of the
CC invention.

XX SQ Sequence 1614 BP; 399 A; 362 C; 493 G; 360 T; 0 U; 0 Other;

Query Match 17.0%; Score 170.8; DB 13; Length 1614;
Best Local Similarity 64.9%; Pred. No. 3e-36;
Matches 253; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 54 CGCCATGGGGAAGTCGGATTCTTACTCCAGGGCTATCGCCAAAGGATCAAGTCCAA 113

Db 76 CACGATGGGGAAGCACGAGTTCTGACGCGAGGCGATCGCAACCGGATCAAGCGAA 135

QY 114 GGGGCTGCAGAGCTACGCTGGTATTGCGCAGATGTCGAGAGCACTGCGGGACGAGAA 173

Db 136 GGGGCTGCAGAGCTGCGGTGGTACTGTCAGATGTCAGAGCACTGCGGGACGAGAA 195

QY 174 TGGCTTTAAGTGTCATTGTATGTCGCAATCTCATCAGACAACTATTGCTGCTTCAGA 233

Db 196 CGGGTTCAAGTGCCTATGATGTCGAGTGCAGATGTCGACGAGCAGATGTCGGCAT 255

QY 234 AAATCCTCAGCAGTTTATGATTTATTTTTCAGAGGAATTCGAAATGACTTTCTAGAAT 293

Db 256 GGGCCCGGACCGGTCGTCGAGGGCTTCTCCGAGGAGTTCCTCGAGTCTCTCCTCCT 315

QY 294 TCTCAGAGACGCTTTGGCAGCTAAAGGGTCCAAACATTTGCTACAAAGATACAT 353

Db 316 CATCCGCGCGGACCGCACCTCCCGCGTCCGCGCCACCGTCTGCTACAAAGATACAT 375

QY 354 CAGCCACCGAGACATCCACATGAATGCTCCTCAGTGGAACTCTGACTGATTTTAC 413

Db 376 CGCCGACCGGACCACTCCATGATGAATCTCCAGCGGTGGCCACCGCTCACCGAGTTCGT 435

QY 414 TAAGTGGCTGGGCGAGAGAGGCTTTGTGCAA 443

Db 436 CAAAGTTCTTGGGCGGAGGGGTACTGTAA 465

RESULT 15

ADX47874

XX ADX47874 standard; cDNA; 1713 BP.

XX AC ADX47874;

XX DX 21-APR-2005 (first entry)

XX Plant full length insert polynucleotide seqid 22614.

DE plant protectant; plant growth regulant; gene therapy; plant;

KW recombinant DNA construct; physical array; plant breeding marker;

KW

KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW extreme osmotic condition; pathogen tolerance; pest tolerance;
KW growth rate; cell cycle pathway; disease resistance;
KW galactomannan production; lignin production; plant growth regulator;
KW yield; plant growth; plant development; seed oil; protein yield;
KW protein content; gene; ss.

XX Unidentified.

XX US2004034888-A1.

PN 19-FEB-2004.

PD 28-APR-2003; 2003US-00425114.

PF 06-MAY-1999; 99US-00304517.

PR 05-NOV-2001; 2001US-00985678.

XX (LIU/) LIU J.

XX (ZHOU/) ZHOU Y.

PA (KOVA/) KOVALIC D K.

PA (SCRE/) SCREEN S E.

PA (TABA/) TABASKA J E.

PA (CAO/) CAO Y.

XX Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;

XX WPI; 2004-180133/17.

XX New recombinant DNA construct, useful for improving plant tolerance to
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT pests, for conferring increased resistance to plant disease, or for
PT improving yield.

PS Claim 1; SEQ ID NO 22614; 15pp; English.

XX The invention describes a recombinant DNA construct comprising a
CC polynucleotide consisting of a sequence encoding an amino acid sequence
CC available in electronic form from the US patent office at
CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC increased resistance to plant disease, for increasing the rate of homologous
CC lignin or plant growth regulators, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This sequence represents a plant full length insert
CC polynucleotide that can be used in the recombinant DNA construct of the
CC invention.

XX SQ Sequence 1713 BP; 423 A; 422 C; 504 G; 364 T; 0 U; 0 Other;

Query Match 17.0%; Score 170.8; DB 13; Length 1713;
Best Local Similarity 64.9%; Pred. No. 3.1e-36;
Matches 253; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 54 CGCCATGGGGAAGTCGGATTCTTACTCCCAAGGCTATCGCCAAAGATCAAGTCCAA 113

Db 235 CACGATGGGGAAGCACGAGTTCTTGACGCGGAAGGCGATCGCAACCGGATCAAGCGAA 294

QY 114 GGGGCTGCAGAGCTACGCTGGTATTGCGAGATGCGCAGAGAGAGTCCGGGACGAGAA 173

Db 295 GGGGCTGCAGAGAGCTGCGGTGGTACTGTCAGATGTCAGAGCAGTCCCGGACGAGAA 354

QY 174 TGGCTTTAAGTGTCATTGTATGTCGGAATCTCATCAGAGACAATATTGCTGGCTTCAGA 233

Db 355 CGGGTTCAAGTGCCTATGATGTCGGATGCGCACGAGGCGAGATGTCAGGTGTTCGGCAT 414

Qy	234	AAATCCTCAGCAGTTTATGANTTATTTTTCAGAGGAATTCGAAATGACTTTCTAGAACT	293
Db	415	GGCCCCCGACCGCGTCGTCGAGGGCTTCTCCGAGGAGTTCTCGAGTCCTTCTCTCCCT	474
Qy	294	TCTCAGGAGACGCTTTTGGCACTAAAAGGGTCCAACAACATTGTCTACAACGAATACAT	353
Db	475	CATCCCGCGCGCACCGCCACTCCCGCGTCGCGCCACCGTCGTCTACAACGAGTACAT	534
Qy	354	CAGCCACCGAGAGCACATCCACATGAATGCCACTCAGTGGGAACTCTGACTGATTTTAC	413
Db	535	CGCCGACCGGCAACAGTCCACATGAATCCACGCGTGGGCCACGCTCACCGAGTTCGT	594
Qy	414	TAAAGTGGCTGGGCAGAGAGGCTTGTGCAA	443
Db	595	CAAGTTCCTGGGGCGGAGGGGTACTGTAA	624

Search completed: November 26, 2005, 13:40:48
Job time : 648.701 secs

This Page Blank (uspic)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 12:58:53 ; Search time 5601.03 Seconds
(without alignments)
8370.011 Million cell updates/sec

Title: US-09-555-529-3

Perfect score: 1002

Sequence: 1 tgattgcgactcggtaccg.....agcatcaaatggtgttcgc 1002

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: gb_est1.*
- 2: gb_est2.*
- 3: gb_est3.*
- 4: gb_est4.*
- 5: gb_est5.*
- 6: gb_est6.*
- 7: gb_est7.*
- 8: gb_est8.*
- 9: gb_gss1.*
- 10: gb_gss2.*
- 11: gb_gss3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	626.4	62.5	1396	4	CR595908 full-leng
2	540.8	54.0	723	3	BQ186497 UI-E-EU1-
3	524.8	52.4	1543	4	AY609991 Sus scrofa
4	510	50.9	1693	7	CN256425 170004245
5	502.2	50.1	1682	1	AV721396 AV721396
6	501	50.0	1182	10	AY412518 Homo sapi
7	501	50.0	1182	10	AY412519 Pan trogl
8	487.4	48.6	733	5	BX104466
9	478	47.7	7848	1	AL558810
10	462.2	46.1	703	1	AM035210
11	462.2	46.1	786	1	AJ819816
12	462.2	46.1	864	1	AM037867
13	455.8	45.5	871	1	AJ819723
14	455.4	45.4	543	6	CB158644 K-EST0218
15	454.6	45.4	720	7	CK301009 UI-E-EU1-
16	451.4	45.0	754	8	DN755964 GL-Cf-127
17	447	44.6	1339	4	CR618602
18	446.2	44.5	800	1	AM038283
19	443.4	44.3	488	3	BM689918 UI-E-CK1-
20	441	44.0	944	1	AL522387
21	437.4	43.7	638	3	BM711163
22	436	43.5	698	1	AI650375 wa90b01.x

C	23	431.4	43.1	701	1	AI807250
C	24	426.4	42.6	820	6	CF409895
C	25	425.2	42.4	949	3	BM453332
C	26	414	41.3	581	3	BP350798
C	27	414	41.3	718	8	CX866441
C	28	414	41.3	882	3	BQ221694
C	29	414	41.3	965	1	AL522388
C	30	413	41.2	552	7	CN256426
C	31	412.6	41.2	908	2	BE894105
C	32	412.4	41.2	922	6	CD389584
C	33	411	41.0	722	1	AM035948
C	34	409.4	40.9	580	3	BQ130747
C	35	408.6	40.8	679	1	AM028238
C	36	404.8	40.4	591	2	BE896845
C	37	402.2	40.1	691	1	AM028686
C	38	400.2	39.9	951	7	CO725903
C	39	400.2	39.9	988	7	CO726806
C	40	397.2	39.6	739	5	BQ782294
C	41	392.4	39.2	424	1	AV721119
C	42	391	39.0	842	3	BI862461
C	43	390	38.9	663	1	AJ660240
C	44	388.8	38.8	776	2	BF571691
C	45	388	38.7	656	7	CV029155

ALIGNMENTS

RESULT 1

CR595908

LOCUS

DEFINITION

full-length cDNA clone CS0DJ015Y19 of T cells (Jurkat cell line)

Cot 10-normalized of Homo sapiens (human).

ACCESSION

CR595908.1

VERSION

HTC; CNSLT_CDNA.

KEYWORDS

Homo sapiens

SOURCE

Homo sapiens

ORGANISM

Hominidae; Homo.

REFERENCE

1. (bases 1 to 1396)

AUTHORS

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

TITLE

Full-length cDNA libraries and normalization

JOURNAL

Unpublished

REMARK

Contact: Feng Liang Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com/

REFERENCE

2. (bases 1 to 1396)

AUTHORS

Genoscope.

TITLE

Direct Submission

JOURNAL

Submitted (20-JUL-2004) Genoscope - Centre National de Sequençage : BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr)

COMMENT

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.

FEATURES

Location/Qualifiers

1. .1396

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0DJ015Y19"

/tissue types="T cells (Jurkat cell line) Cot 10-normalized"

/plasmid="pCMVSPORT_6"

ORIGIN

Query Match

Best Local Similarity

Matches

942; Conservative

76.1%;

Pred. No. 4.3e-151;

Indels

295; Gaps

2;

library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI). "

ORIGIN	Query Match	54.0%;	Score 540.8;	DB 3;	Length 723;
	Best Local Similarity	99.1%;	Pred. No. 5.9e-129;		
	Matches 553;	Conservative 0;	Mismatches 4;	Indels 1;	Gaps 1;
QY	445	AGTGCACCTGAAGACGATAGGAGTTTCAGCATCTAGTGAACCGAAAGAAATCTTCCAGAGC	504		
DB	39	AGTGCACCTGAAGACGATAGGAGTTTCAGCATCTAGTGAACCGAAAGAAATCTTCCAGAGC	98		
QY	505	TCAACTCAGTCTTAAAGAAAAAGAAAAAGAAATCTGCACCTGGATGAAATCATGGAGATT	564		
DB	99	TCAACTCAGTCTTAAAGAAAAAGAAAAAGAAATCTGCACCTGGATGAAATCATGGAGATT	158		
QY	565	GAAGAGAAAAAGAAAAAGAACTGCCCGCAACAGACTACTGGCTACAGCTGAATTTATTGTG	624		
DB	159	GAAGAGAAAAAGAAAAAGAACTGCCCGCAACAGACTACTGTCTACAGCTGAATTTATTGTG	218		
QY	625	AAAATTATACCAAGAAACTGGGAGAGAAATATCATAGAAAAAGGCTATTGTTAAGGAA	684		
DB	219	AAAATTATACCAAGAAACTGGGAGAGAAATATCATAGAAAAAGGCTATTGTTAAGGAA	278		
QY	685	GTAATTGACAAATATACAGCTGTTTGTGAAGATGATTGATCTGGAGACAAGCTGAAACTT	744		
DB	279	GTAATTGACAAATATACAGCTGTTTGTGAAGATGATTGATCTGGAGACAAGCTGAAACTT	338		
QY	745	GACCAGACTCAATTTAGAGACAGTAAATTCACAGCACGAGAAAGAAATCTAGTGTTTAAAT	804		
DB	339	GACCAGACTCAATTTAGAGACAGTAAATTCACAGCACGAGAAAGAAATCTAGTGTTTAAAT	398		
QY	805	GGAGGCTACAGAGGAAATGAAGGTACCCCTAGAAATCCATCAATGAGAAAGACTTTTCAGCT	864		
DB	399	GGAGGCTACAGAGGAAATGAAGGTACCCCTAGAAATCCATCAATGAGAAAGACTTTTCAGCT	458		
QY	865	ACTATCGTCATGTGAAACTGGGCCCTTTAAAGGACGCGAGGTTGAAGGAATTCATATGAA	924		
DB	459	ACTATCGTCATGTGAAACTGGGCCCTTTAAAGGACGCGAGGTTGAAGGAATTCATATGAA	518		
QY	925	GACATTTCTAAACTTGGCTGAGTTTGAAATTTGTTAAACATACCTTTAAAAATCTTAAAG	984		
DB	519	GACATTTCTAAACTTGGCTGAGTTTGAAATTTGTTAAACATACCTTTAAAAATCTTAAAG	577		
QY	985	CATCAAAATTTGGTTCGC	1002		
DB	578	CATCANATTTGGTTCGC	595		

FEATURES	50, Tjele DK-8830, Denmark
source	Location/Qualifiers
	1. .1543
	/organism="Sus scrofa"
	/mol_type="mrna"
	/db_xref="taxon:9823"
	/clone="CLU_5302.scr.mek.pl.Contig4"
misc_feature	1. .1543
	/notes="similar to NM_012311.2 Homo sapiens KIN, antigenic determinant of recA protein homolog (mouse) (KIN)"
ORIGIN	
Query Match	52.4%; Score 524.8; DB 4; Length 1543;
Best Local Similarity	70.7%; Pred. No. 9.9e-125;
Matches 880; Conservative	0; Mismatches 72; Indels 292; Gaps 2;
Qy	47 CCGTGGTCGCATGCGGAAGCTCGGATTTTCTTACTCCCAAGGCTATCGCCAAACAGGATCA 105
Db	
Qy	25 CTGTGGTCGCATGCGGAAGTCCGATTTTCTTACCCCGAAGGCCATCGCCAAACAGGATCA 84
Db	
Qy	107 AGTCCAAGGGCTCCAGAAAGCTACGTGGTATTGCGCAGATGTGCCAAGACGAGTGCCTGG 166
Db	85 AATCCAAGGGCTTCAGAAGCTCGCTGGTATTGCGCAGATGTGCCAAGACGAGTGCCTGG 144
Qy	167 ACGAAGATGCTTTAAGTGTCAATGTATGTCCGAATCTCATCAGACACAACTATTGCTGG 226
Db	145 ACGAAGATGCTTTAAGTGTCAATGTATGTCCGAATCGCATCAGACAACTATTGCTGG 204
Qy	227 CTTTCAGAAAATCCTCAGCAGCTTTATGATGATTTTTCAGAGGAATTCGGAATGACTTTTC 286
Db	
Qy	205 CTTCTGAAAATCCTCAGCAGTTTATGATTTATGATTTTTCAGAGGAATTCGGAATGACTTTTC 264
Qy	287 TAGAACTTCTCAGGAGACGCTTTTGGCACTAAAAAGGGTCCAACAACATTTGTCTACAACG 346
Db	265 TAGAACTTCTCAGGAGACGCTTTTGGCACCAGAGGGTTTCAACAACATCGTCTACAACG 324
Qy	347 AATACATCAGCCACCGAGAGACATCCACATGAATGCACTCAGTGGGAACCTCTGACTG 406
Db	325 AGTACATCAGCCACCGAGAGACATCCACATGAACGCCACGACGTGGGAGACGCTGACGG 384
Qy	407 ATTTTACTAGTCGCTGGGCAGAGAGGCTTGTGCA- - - - - 443
Db	
Qy	385 ATTTTCAACAAGTGGCTGGGCAGAGAGGCTTGTGCAAAAGTGATGAGACACCAAAAGGCT 444
Qy	444 - - - - - 443
Db	445 GGTATATTAGTACATAGATAGGACCCGGAAACAATCCGCGGCAACTGGAACCTAGAGA 504
Qy	444 - - - - - 443
Db	505 AAAAGAAGAAGCAGGACTTGACGATGAAGAAAAAACTGCCAAATTTATTCAAGAAACAAG 564
Qy	444 - - - - - 443
Db	565 TGAGACGAGGTCTGGAAGGGAAGGAGCAGGAGGCCCTATTTTTCAGGAGTTAAGCAGAG 624
Qy	444 - - - - - 443
Db	625 AAAATGAAGAAGAAAAAGTGACATTTAATTTGAACAAAGGAGCATGTAGTTCAGCAGCAG 684
Qy	444 - - - - - 443
Db	685 CATCTTCCAAATCAAGTTCTTTTGGGACCAAGTGCCTTGAGACGATGGGTAGCAGCAGCGT 744
Qy	476 CAGTGAACCGAAAAGATCTTCCAGAGCTCAACTCAGTCTTAAAGAAAAGAGAAAAGA 535
Db	745 CGGTGAACCGAAGGAATCTTCCAGAGCTCAGCTCAAAAGAAAAGAAAAGA 804
Qy	536 AATCTGCACCTGGATGAATCATGAGATTTGAAGAGGAAGAAAAGAACTGCCCGCAACAG 595
Db	805 AGTCTGCGCTCGAAGAAATCATGAGATTTGAAGAGGAGAGAGAGAACTGCCCGGACGG 864
Qy	596 ACTACTGGCTACAGCCTGAAATTTATGTGAAAAATTTATAACCAAGAAACTGGGAGAGAAAT 655

Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
Science 302 (5652), 1960-1963 (2003)
14671302
2 (bases 1 to 1182)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civallo,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
Location/Qualifiers
Source
1..1182
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>1182
/gene="KIN"
/locus_tag="HCM4584"
ORIGIN
Query Match 50.0%; Score 501; DB 10; Length 1182;
Best Local Similarity 100.0%; Pred. No. 1.4e-118;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 445 AGTGCACGATAGAGGAGTTCAGCATGTCAGTAAAGCAAGAAATCTCCAGAGC 504
DB 682 AGTGCACGATAGAGGAGTTCAGCATGTCAGTAAAGCAAGAAATCTCCAGAGC 741
QY 505 TCAACTCAGCTAAAGAAAAGAAAGAAATCTGCATGCTGATGAAATCATGGAGATT 564
DB 742 TCACTCAGCTAAAGAAAAGAAAGAAATCTGCATGCTGATGAAATCATGGAGATT 801
QY 565 GAAGAGGAAAAGAAAAGAAATCTGCCGAAACAGACTACTGGCTACAGCCTGAAATTTATTGTG 624
DB 802 GAAGAGGAAAAGAAAAGAAATCTGCCGAAACAGACTACTGGCTACAGCCTGAAATTTATTGTG 861
QY 625 AAAAAATTAACCAAGAAATCGGAGAGAAATATCATAGAAAAGCTATTGTTAAGGAA 684
DB 862 AAAAAATTAACCAAGAAATCGGAGAGAAATATCATAGAAAAGCTATTGTTAAGGAA 921
QY 685 GTAATTGACAAATATACAGCTGTTGTGAAGATGATTGATTTCTGGAGACAAGCTGAAACTT 744
DB 922 GTAATTGACAAATATACAGCTGTTGTGAAGATGATTGATTTCTGGAGACAAGCTGAAACTT 981
QY 745 GACCAAGACTCATTTAGAGACAGTAATTCAGCACCGAGAAAGAAATTTAGTTTAAAT 804
DB 982 GACCAAGACTCATTTAGAGACAGTAATTCAGCACCGAGAAAGAAATTTAGTTTAAAT 1041
QY 805 GGAGGCTACAGAGGAATGAAGTACCTTAGAATCCATCAATGAGAGACTTTTTCAGCT 864
DB 1042 GGAGGCTACAGAGGAATGAAGTACCTTAGAATCCATCAATGAGAGACTTTTTCAGCT 1101
QY 865 ACTATCGTCATTGAAACTGCGCCCTTTAAAGGAGCGAGAGTTGAAAGCAATTCATATGAA 924
DB 1102 ACTATCGTCATTGAAACTGCGCCCTTTAAAGGAGCGAGAGTTGAAAGCAATTCATATGAA 1161
QY 925 GACATTTCTAAACTTGCCTGA 945
DB 1162 GACATTTCTAAACTTGCCTGA 1182
RESULT 8
BX104466/c
LOCUS
DEFINITION
BX104466 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
IMAGE:230344; IMAGE:230344, mRNA sequence.
BX104466
,ACCESSION

VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
BX104466.1 GI:27832949
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo
1 (bases 1 to 733)
Ebert,L., Heil,O., Hennig,S., Neubert,P., Partsch,E., Peters,M.,
Radelof,U., Schneider,D. and Korn,B.
Human Unigeneset - RZPD3
Unpublished (2003)
Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
RZPD; IMAGE:230344
RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
Human Unigeneset - RZPD3 (RZPDLIB No.972)
http://www.rzpd.de/CloneCards/cgi-
bin/showLib.pl.cgi/response?libNo=972 Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 101
Fax: +49 30 32639 111
www.rzpd.de
This clone is available royalty-free from RZPD;
contact RZPD (clone@rzpd.de) for further information. Seq primer:
M13r, Primer sequence: TTTCCACAGAAAACAGCTATGAC.
FEATURES
Location/Qualifiers
Source
1..733
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:230344"
/sex="male"
/lab_host="Soares fetal liver spleen INFLS"
/dev_stage="20 week-post conception fetus"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site.1: Pac I; Site.2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACGGAAGAAATTAATGAAGATCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Patima Bonaldo."
ORIGIN
Query Match 48.6%; Score 487.4; DB 5; Length 733;
Best Local Similarity 99.6%; Pred. No. 4.2e-115;
Matches 499; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 502 AGCTCAACTCAGCTCTAAAGAAAAGAAAGAAATCTGCACCTGATGAAATCATGGAG 561
DB 733 AGCTCAACTCAGCTCTAAAGAAAAGAAAGAAATCTGCACCTGATGAAATCATGGAG 674
QY 562 ATTGAAGAGGAAAAGAAAGAAATCTGCCGAAACAGACTACTGGCTACAGCCTGAAATTTATT 621
DB 673 ATTGAAGAGGAAAAGAAAGAAATCTGCCGAAACAGACTACTGGCTACAGCCTGAAATTTATT 614
QY 622 GTGAAAATTTATACCAAGAAACTGGGAGAGAAATATCATAGAAAAGCTATTGTTAAG 681
DB 613 GTGAAAATTTATACCAAGAAACTGGGAGAGAAATATCATAGAAAAGCTATTGTTAAG 554
QY 682 GAAGTAATTGACAAATATACAGCTGTTGTGAAGATGATTGATTTCTGGAGACAAGCTGAAA 741
DB 553 GAAGTAATTGACAAATATACAGCTGTTGTGAAGATGATTGATTTCTGGAGACAAGCTGAAA 494
QY 742 CTTGACCAGACTCATTTAGAGACAGTAATTCAGCACCGAGAAAGAAATTTAGTTTAA 801
DB 493 CTTGACCAGACTCATTTAGAGACAGTAATTCAGCACCGAGAAAGAAATTTAGTTTAA 434

```

QY 802 AATGAGGCTACAGAGAAATGAAGTACCGCTAGATCCATCAATGAGAGAACTTTTCA 861
DB 433 AATGAGGCTACAGAGAAATGAAGTACCGCTAGATCCATCAATGAGAGAACTTTTCA 374
QY 862 GCTACTATCGTCAATGAAGTCCCTTTAAAGAGCGAGAGTGAAGGAATTTCAATAT 921
DB 373 GCTACTATCGTCAATGAAGTCCCTTTAAAGAGCGAGAGTGAAGGAATTTCAATAT 314
QY 922 GAAGACATTTCAACTTCGCTGAGTTGAAATTTGTTAAACATACCTTTAAATCTTA 981
DB 313 GAAGACATTTCAACTTCGCTGAGTTGAAATTTGTTAAACATTA-CATTAAATCTTA 255
QY 982 AAGCATCAATTTGTTGCTGC 1002
DB 254 AAGCATCAATTTGTTGCTGC 234

RESULT 9
AL558810/c 848 bp mRNA linear EST 02-APR-2004
LOCUS AL558810 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
DEFINITION Homo sapiens cDNA clone CS0DJ015Y119 3-PRIME, mRNA sequence.
ACCESSION AL558810
VERSION AL558810.2 GI:31280608
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 848)
AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On Feb 15, 2001 this sequence version replaced gi:12903696.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
5543.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0DJ015AE10NP1&c=5543.r.
FEATURES
source
Location/Qualifiers
1..848
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DJ015Y119"
/cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
/cell_line="JURKAT"
/clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT
10-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and EcoR V into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Query Match 47.7%; Score 478; DB 1; Length 848;
Best Local Similarity 89.1%; Pred. No. 1.2e-112;
Matches 497; Conservative 21; Mismatches 39; Indels 1; Gaps 1;
QY 445 AGTCACATGAAGCATGAGTTCAGCATCAGTGAACGAAAGAAATCTTCCAGC 504
DB 720 AGBCACKAGACGADGAGGAAGTTCAGCAWCAKGAACGAAAGAAAGCCCTGNGC 661
QY 505 TCACTCAGTCTAAGAAAGAAAGAAATCTGCCTGATGAATCATCGAGATT 564

```

```

DB 660 TCATWTTTTTTTAAAGTATTTTGAAGAAATCTGCACTGGATGAATCANGGAGCC 601
QY 565 GAAGAGGAAAGAAAGAACTGCGCGAACAGACTACTGGCTACAGCTCTGAAATTTATGTG 624
DB 600 GAAGAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 541
QY 625 AAAATTTATCAACCAAGAACTGGGAGAGAAATATATATAAGAAAGAAAGAAAGAAAG 684
DB 540 AAAATTTATCAACCAAGAACTGGGAGAGAAATATATATAAGAAAGAAAGAAAGAAAG 481
QY 685 GTAATGCAAAATATACAGCTGTTGTGAAGATGATTTCTGGAGACAGCTGAAACTT 744
DB 480 GTAATGCAAAATATACAGCTGTTGTGAAGATGATTTCTGGAGACAGCTGAAACTT 421
QY 745 GACCGACTCATTTAGAGACAGTAATTCAGACACAGGAAAGAAAGAAATTTAGTTTAAAT 804
DB 420 GACCGACTCATTTAGAGACAGTAATTCAGACACAGGAAAGAAAGAAATTTAGTTTAAAT 361
QY 805 GGAGGCTACAGAGGAAATGAAGTACCTAGATCCATCAATGAGAGAACTTTTTCAGCT 864
DB 360 GGAGGCTACAGAGGAAATGAAGTACCTAGATCCATCAATGAGAGAACTTTTTCAGCT 301
QY 865 ACTATCGTCAATGAAGTCCCTTTAAAGAGCGAGAGTGAAGAAATTTCAATATGAA 924
DB 300 ACAATCGCAAGGAGACGCGCCCGGAAAGGAGCGAGAGTTGAAGGAAATTTCAATATGAA 241
QY 925 GACATTTCTAAATTCGCTGAGTTTGAAATTTGTTAAACATACCTTTTAAATCTTAAAG 984
DB 240 GACATTTCTAAATTCGCTGAGTTTGAAATTTGTTAAACATTA-CATTAAATCTTAAAG 182
QY 985 CATCAATTTGTTGCTGC 1002
DB 181 CATCAATTTGTTGCTGC 164

RESULT 10
AL558810/c 703 bp mRNA linear EST 13-JUL-2005
LOCUS AL558810 KN-252-spleen, Bos taurus Bos taurus cDNA clone
DEFINITION C000740112 5', mRNA sequence.
ACCESSION AL558810
VERSION AL558810.1 GI:70790518
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
REFERENCE 1 (bases 1 to 703)
AUTHORS Anderson, S.I., Taylor, R., Talbot, R., Speed, D., Law, A.S.,
Humphray, S., Hanotte, O., Mwakaya, J. and Archibald, A.L.
TITLE Development of cDNA and EST resources for studying host responses
to trypanosome infection in cattle
JOURNAL Unpublished (2005)
COMMENT Contact: Anderson SI
Genomics and Bioinformatics
Roslin Institute
Roslin, Midlothian, EH25 9PS, UNITED KINGDOM.
FEATURES
source
Location/Qualifiers
1..703
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="C000740112"
/tissue_type="spleen"
/clone_lib="KN-252-spleen, Bos taurus"
/notes="breed: N'dama"
ORIGIN
Query Match 46.1%; Score 462.2; DB 1; Length 703;
Best Local Similarity 89.6%; Pred. No. 1.4e-108;
Matches 508; Conservative 0; Mismatches 58; Indels 1; Gaps 1;

```

QY 436 TTGTGCAAAAGTGCATGAGCGATAGGAGTTTCAGCATCAGTGAACGAAAGAAATCT 495
Db 133 TTGGGACCAAGTGCCTGAAACGATTTGGGACCGTGGCATCAGTGAAGCGGAGGAGTCC 192
QY 496 TCCGAGAGCTCAACTCAGTCTAAAGAAAAAGAAAAAGAAATCTGCACTGGATGAAATC 555
Db 193 TCCGAGAGCTCGGCTCCGTCMAAGGAGAGAGAGAAAGTCTGCCCTCGATGAATC 252
QY 556 ATGGAGATTGAAGAGAGAGAAAAAGAACTGCCCGAACAGACTACTGGCTACAGCCTGAA 615
Db 253 ATGGAGATTGAAGAGAGAGAAAAAGAACTGCCCGAACAGACTACTGGCTGAGCCTGAA 312
QY 616 ATTATTGTGAAATATATACCAAGAACTGGGAGAGAAATATCATAGAAAAGGCTATT 675
Db 313 ATCATAGTGAATATATACCAAAAACTTTGGAGAGAGTATCATAGAAAGAGGGGATT 372
QY 676 GTTAAGGAAGTAAATTCACAAATATACAGCTGTGTGTGAAGATGATTGATCTCGAGACAAG 735
Db 373 GTTAAGGAAGTAAATTCACAAATATACAGCTGTGTGAAGATGATTGATCTCGAGACAAG 432
QY 736 CTGAAACTTGCACAGACTCATTTAGAGACAGTAATTTCCAGCCAGGAAAAAGAAATCTTA 795
Db 433 CTGAAACTTGCACAGACTCATTTAGAGACAGTAATTTCCAGCCAGGAAAAAGAAATCTTT 492
QY 796 GTTTTAAATGGAGGCTACAGAGAAATGAAGTACCTAGATCCATCAATGAGAGACT 855
Db 493 GTTTTAAATGGAGGCTACAGAGAAATGAAGGACCTTGAATCCATCAATGAGAGACT 552
QY 856 TTTTCAGCTACTATCGTCAATGAAACTGGCCCTTTTAAAGGCGCAGAGTTTGAAGGAAT 915
Db 553 TTTTCGCTACTATAGTCAATGAACTGGCCCTTTTAAAGGCGCAGAGTTTGAAGGAAT 612
QY 916 CAATATGAAGACATTTCTAACTGCTGAGTTTGAAGAAATTTGTTAAACAATACCTTTAAA 975
Db 613 CAATACGAAGACATTTCTAACTGCTGAGTTTGAAGAAATTTGTTAGC-ATCACATTA 671
QY 976 ATCTTAAAGCATCAAAATGGTGTCCG 1002
Db 672 ATCTTAAAGCATCAAAATGGTGTTCAC 698

RESULT 11
AJ819816
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source

AJ819816 786 bp mRNA linear EST 12-MAY-2005
AJ819816 KN206 Bos sp. cDNA clone C0006016j08, mRNA sequence.
AJ819816.1 GI:51887292
EST.
Bos sp.
Bos sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 786)
McGuire,K. and Glass,E.J.
The expanding role of microarrays in the investigation of
macrophage responses to pathogens
Vet. Immunol. Immunopathol. 105 (3-4), 259-275 (2005)
Contact: McGuire K
Genomics and Genetics
Roslin Institute
Roslin, Midlothian, EH25 9PS, UNITED KINGDOM
Single pass sequencing. Bases called and trimmed with phred
v0.020425.c. Vector identified by cross match with the -minscore 20
and -minmatch 12 options. Vector:pBluescriptII(SK+) R. Site
1:EcoRV(lost) R. Site 2:NotI Seq Primer: T7 Normalised library
constructed from pooled monocytes from Bos taurus (Holstein) and
Bos indicus (Sahiwal) cattle subjected to various stimuli,
including infection with the protozoan parasite Theileria annulata.
Location/Qualifiers
1. .786
/organism="Bos sp."
/mol_type="mRNA"

/db_xref="taxon:29061"
/clone="C0006016j08"
/tissue_type="blood"
/cell_type="bovine monocyte"
/clone_lib="KN206"
/note="Vector: pBluescriptII(SK+); Site 1: EcoRV(lost);
Site 2: NotI; Normalised library constructed from pooled
monocytes from Bos taurus (Holstein) and Bos indicus
(Sahiwal) cattle subjected to various stimuli, including
infection with the protozoan parasite Theileria annulata"

Query Match 46.1%; Score 462.2; DB 1; Length 786;
Best Local Similarity 89.6%; Pred. No. 1.5e-108;
Matches 508; Conservative 0; Mismatches 58; Indels 1; Gaps 1;

ORIGIN

QY 436 TTGTGCAAAAGTGCATGAGCGATAGGAGTTTCAGCATCAGTGAACGAAAGAAATCT 495
Db 204 TTGGGACCAAGTGCCTGAAACGATTTGGGACCGTGGCATCAGTGAAGCGGAGGAGTCC 263
QY 496 TCCGAGAGCTCAACTCAGTCTAAAGAAAAAGAAAAAGAAATCTGCACTGGATGAAATC 555
Db 264 TCCGAGAGCTCGGCTCCGTCMAAGGAGAGAGAAAGTCTGCCCTCGATGAAATC 323
QY 556 ATGGAGATTGAAGAGAGAGAAAAAGAACTGCCCGAACAGACTACTGGCTACAGCCTGAA 615
Db 324 ATGGAGATTGAAGAGAGAGAAAAAGAACTGCCCGAACAGACTACTGGCTGAGCCTGAA 383
QY 616 ATTATTGTGAAATATATACCAAGAACTGGGAGAGAAATATCATAGAAAAGGCTATT 675
Db 384 ATCATAGTGAATATATACCAAAAACTTTGGAGAGAGTATCATAGAAAGAGGGGATT 443
QY 676 GTTAAGGAAGTAAATTCACAAATATACAGCTGTGTGTGAAGATGATTGATCTCGAGACAAG 735
Db 444 GTAAAGGAAGTAAATTCACAAATATACAGCTGTGTGTGAAGATGATTGATCTCGAGACAAG 503
QY 736 CTGAAACTTGCACAGACTCATTTAGAGACAGTAATTTCCAGCCAGGAAAAAGAAATCTTA 795
Db 504 CTGAAACTTGCACAGACTCATTTAGAGACAGTAATTTCCAGCCAGGAAAAAGAAATCTTT 563
QY 796 GTTTTAAATGGAGGCTACAGAGGAAATGAAGTACCTTAGAATCCATCAATGAGAGACT 855
Db 564 GTTTTAAATGGAGGCTACAGAGGAAATGAAGCACCCTTAGAATCCATCAATGAGAGACT 623
QY 856 TTTTCAGCTACTATCGTCAATGAACTGGCCCTTTTAAAGGCGCAGAGTTTGAAGGAAT 915
Db 624 TTTTCGCTACTATAGTCAATTTGAATCTGCCCTTTTAAAGGCGCAGAGTTTGAAGGAAT 683
QY 916 CAATATGAAGACATTTCTAACTGCTGAGTTTGAAGAAATTTGTTAAACAATACCTTTAAA 975
Db 684 CAATACGAAGACATTTCTAACTGCTGAGTTTGAAGAAATTTGTTAGC-ATCACATTA 742
QY 976 ATCTTAAAGCATCAAAATGGTGTTCGC 1002
Db 743 ATCTTAAAGCATCAAAATGGTGTTCAC 769

RESULT 12
AM037867
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

AM037867 864 bp mRNA linear EST 13-JUL-2005
C0007409b03 5', mRNA sequence.
AM037867
AM037867.1 GI:70793175
EST.
Bos taurus (cow)
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 864)
Anderson,S.I., Taylor,R., Talbot,R., Speed,D., Law,A.S.,
Humphray,S., Hanotte,O., Wakaya,J. and Archibald,A.L.


```

QY 976 ATCTTAAAGCATCAAAATTCGGTGTTCGC 1002
| | | | | | | | | | | | | | | | | |
Db 152 ATCTTAAAGCATCAAAATTCGGTGTTCAC 126
| | | | | | | | | | | | | | | | | |

RESULT 14
Cb158644
LOCUS K-EST0218017 L18POOLin1 543 bp mRNA linear EST 29-JAN-2003
DEFINITION 5', mRNA sequence.
ACCESSION Cb158644
VERSION Cb158644.1 GI:28143780
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 543)
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 12 row: F column: 09
High quality sequence stop: 543.

FEATURES
source
1..543
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L18POOLin1-12-F09"
/cell_line="SNU-354+Cho-CK+Choi-CK+HLK-3"
/lab_host="Top10P"
/clone_lib="L18POOLin1"
/notes="Organ: Liver; Vector: pTT3-Pac; Site 1: EcoRI;
Site 2: NotI; The library was contributed by the Soares
laboratory and it was constructed as described by Bonaldo,
M.F., Lennon, G. and Soares, M.B. (1996), Genome Research
6(9): 791-806. RNA was prepared from harvested cell
culture."

ORIGIN
Query Match 45.4%; Score 455.4; DB 6; Length 543;
Best Local Similarity 99.6%; Pred. No. 7.8e-107;
Matches 467; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 534 GAAATCTGCATGGAATCATGAGATTGAAGAGGAAAAGAAAGAACTGCCCGAAC 593
| | | | | | | | | | | | | | | | | |
Db 1 GAAATCTGCATGGAATCATGAGATTGAAGAGGAAAAGAAAGAACTGCCCGAAC 60
| | | | | | | | | | | | | | | | | |

QY 594 AGACTACTGGCTACACCTGAAATTTATGTAATTTATACCAAGAACTGGAGAGAA 653
| | | | | | | | | | | | | | | | | |
Db 61 AGACTACTGGCTACACCTGAAATTTATGTAATTTATACCAAGAACTGGAGAGAA 120
| | | | | | | | | | | | | | | | | |

QY 654 ATATCATAGAAAAGCGTATTGTTAAGGAAGTAAATTGACAAATATACAGCTGTTGTGAA 713
| | | | | | | | | | | | | | | | | |
Db 121 ATATCATAGAAAAGCGTATTGTTAAGGAAGTAAATTGACAAATATACAGCTGTTGTGAA 180
| | | | | | | | | | | | | | | | | |

QY 714 GATGATTGATTCTGGAGACAAGCTGAAACTTGACAGACTCAATTTAGACAGAGTAATCC 773
| | | | | | | | | | | | | | | | | |
Db 181 GATGATTGATTCTGGAGACAAGCTGAAACTTGACAGACTCAATTTAGACAGAGTAATCC 240
| | | | | | | | | | | | | | | | | |

QY 774 AGCACCAGAAAAGATTCCTAGTTTAAATGCGGCTACAGGGAATGAGGTACCCCT 833
| | | | | | | | | | | | | | | | | |

```

```

Db 241 AGCACCAGAAAAGATTCCTAGTTTAAATGCGGCTACAGGGAATGAGGTACCCCT 300
| | | | | | | | | | | | | | | | | |
QY 834 AGAATCCATCAATGAGAAGACTTTTTCAGCTACTATCTCATTTGAAACTGGCCCTTTAAA 893
| | | | | | | | | | | | | | | | | |
Db 301 AGAATCCATCAATGAGAAGACTTTTTCAGCTACTATCTCATTTGAAACTGGCCCTTTAAA 360
| | | | | | | | | | | | | | | | | |

QY 894 AGGAGCGAGCTTGAAGGAATTCATATGAAGACATTTCTAAACTTGCCTGAGTTTGAAA 953
| | | | | | | | | | | | | | | | | |
Db 361 AGGAGCGAGCTTGAAGGAATTCATATGAAGACATTTCTAAACTTGCCTGAGTTTGAAA 420
| | | | | | | | | | | | | | | | | |

QY 954 ATTTGTTAAACAATACCTTTAAAATCTTAAAGCATCAAAATTCGTGTTCGC 1002
| | | | | | | | | | | | | | | | | |
Db 421 ATTTGTTAAACAAT-CATTAAAATCTTAAAGCATCAAAATTCGTGTTCGC 468
| | | | | | | | | | | | | | | | | |

RESULT 15
CK301009/c
LOCUS UI-E-EJ1-ajr-g-04-0-UI.s1 720 bp mRNA linear EST 15-DEC-2003
DEFINITION UI-E-EJ1-ajr-g-04-0-UI 3', mRNA sequence.
ACCESSION CK301009
VERSION CK301009.1 GI:39890957
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 720)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
PUBMED 8889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Greg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/eye.html
Seq primer: M13 FORWARD
POLYA=Yes.
Location/Qualifiers
1..720
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-EJ1-ajr-g-04-0-UI"
/tissue_type="fetal eyes, lens, eye anterior segment,
optic nerve, retina, Retina Foveal and Macular, RPE and
Choroid"
/dev_stages="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-EJ1"
/notes="Organ: eye; Vector: pTT3-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-E-EJ1 is a subtracted cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pTT3-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are: fetal eyes,

```


AGAATCAAGA: lens, CGATTAGCGA: eye anterior segment, AATGCCGAT; optic nerve, CCATTAGTG; retina, CCGCG; Retina Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI). TAG TISSUE=human lens

TAG_LIB=UI-E-EJ1

TAG_SEQ=CGATTAGCGA"

ORIGIN

Query Match	45.4%	Score	454.6	DB	7	Length	720		
Best Local Similarity	98.9%	Pred. No.	1.3e-106						
Matches	468	Conservative	0	Mismatches	4	Indels	1	Gaps	1
Qy	530	AAAAAATCTGCTGATGAAATCATGAGATTGAGAGGAAAGAAAAGAACTGCC	589						
Db	715	AAAAAATCTGCTGATGAAATCATGAGATTGAGAGGAAAGAAAAGAACTGCC	656						
Qy	590	GAACAGACTCTGCTACAGCTGAAATTTTGTGAAAATTTATACCAAGAACTGGGAG	649						
Db	655	GAACAGACTCTGCTACAGCTGAAATTTTGTGAAAATTTATACCAAGAACTGGGAG	596						
Qy	650	AGAAATATCATAGAAAAGGCTATTCTTAAGGAAGTAAATGACAAATATACAGCTGTG	709						
Db	595	AGAAATATCATAGAAAAGGCTATTCTTAAGGAAGTAAATGACAAATATACAGCTGTG	536						
Qy	710	TGAAGATGATTCTTGGAGACAAGCTGAAAATTGACCAAGACTCATTAGAGACAGTAA	769						
Db	535	TGAAGATGATTCTTGGAGACAAGCTGAAAATTGACCAAGACTCATTAGAGACAGTAA	476						
Qy	770	TTCAGCACCAGGAAAGAAATCTAGTTTTAAATGGAGCTACAGAGGAAATGAAGGTA	829						
Db	475	TTCAGCACCAGGAAAGAAATCTAGTTTTAAATGGAGCTACAGAGGAAATGAAGGTA	416						
Qy	830	CCCTAGAAATCCATCAATGAGAAGCTTTTTCAGCTACTATCGTCAATTGAAACTGGCCCTT	889						
Db	415	CCCTAGAAATCCATCAATGAGAAGCTTTTTCAGCTACTATCGTCAATTGAAACTGGCCCTT	356						
Qy	890	TAAAGGACGAGAGTTGAGGAAATTCATATGAAGACATTTCTAAACTTGCCTGAGTTT	949						
Db	355	TAAAGGACGAGAGTTGAGGAAATTCATATGAAGACATTTCTAAACTTGCCTGAGTTT	296						
Qy	950	GAATTTTGTAAACAATACCTTTAAATCTTAAAGCATCAAAATGGTTCGC	1002						
Db	295	GAATTTTGTAAACAATACCTTTAAATCTTAAAGCATCAAAATGGTTCGC	244						

Search completed: November 27, 2005, 00:57:50
Job time : 5606.03 secs

This Page Blank (uspjc)

Result No.	Score	Query Match	Length	DB	ID	Description	
C 1	140	14.0	23645	3	US-09-949-016-13916	Sequence 13916, A	
C 2	123.6	12.3	538	3	US-09-270-767-379	Sequence 379, App	
C 3	123.6	12.3	538	3	US-09-270-767-15661	Sequence 15661, A	
C 4	79	7.9	825	3	US-09-248-796A-5938	Sequence 5938, Ap	
C 5	74.6	7.4	7218	2	US-08-232-463-14	Sequence 14, Appl	
C 6	44.6	4.5	137394	3	US-09-949-016-13872	Sequence 13872, A	
C 7	44.6	4.5	137743	3	US-09-949-016-12178	Sequence 12178, A	
C 8	44.2	4.4	1832	3	US-09-621-976-2813	Sequence 2813, Ap	
C 9	43.6	4.4	612	3	US-09-302-540-1357	Sequence 1357, Ap	
C 10	42.2	4.2	168174	3	US-10-071-411A-63	Sequence 63, Appl	
C 11	42.2	4.2	188273	3	US-10-071-411A-2	Sequence 2, Appli	
C 12	42	4.2	601	3	US-09-949-016-89871	Sequence 89871, A	
C 13	42	4.2	601	3	US-09-949-016-162984	Sequence 162984,	
C 14	41.6	4.2	2518	3	US-09-433-699-3	Sequence 3, Appli	
C 15	41.6	4.2	2527	3	US-09-949-016-4169	Sequence 4169, Ap	
C 16	41.6	4.2	2637	3	US-10-104-047-681	Sequence 681, Appl	
C 17	41.4	4.1	202001	3	US-09-734-674-3	Sequence 3, Appli	
C 18	41.4	4.1	202001	3	US-10-274-990-3	Sequence 3, Appli	
C 19	41.2	4.1	601	3	US-09-949-016-93431	Sequence 93431, A	
C 20	41.2	4.1	96987	3	US-09-949-016-14429	Sequence 14429, A	
C 21	41	4.1	41736	3	US-09-949-016-17091	Sequence 17091, A	
C 22	40.8	4.1	601	3	US-09-949-016-89870	Sequence 89870, A	
C 23	40.8	4.1	601	3	US-09-949-016-162983	Sequence 162983,	
C 24	40.8	4.1	33109	3	US-09-949-016-14301	Sequence 14301, A	

; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 379
; LENGTH: 538
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-379

Query Match 12.3%; Score 123.6; DB 3; Length 538;
Best Local Similarity 57.5%; Pred. No. 2.2e-26;
Matches 222; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

QY 553 ATCATGGAGATTGAAGAGGAAAGAAAGAACTGCCCGAACACAGACTACTGGCTACAGCCT 612
DB 522 ATCAAGCAGGAGAAAGCAAAAGGAGCGTGCACCGCAAAAGACTACTGGCTGCACAA 463
QY 613 GAAATATTGTGAAATATTAAACCAAGAACTGGGAGAGAAATATCATAGAAAAAGGCT 672
DB 462 GGTATCGTGTCAATTTATTTCCAAATCCATGGGCGAAAGTTCTTCAACAAAAAGCG 403
QY 673 ATGTGTAAGAAAGTAATTGACAAATATACAGCTGTTGTGAAGATGATGATCTTGAGAC 732
DB 402 GTTGTCTCGAGCTAATTGACAGATATCAGGCGCAAAATCAAGTTCTTGAGAGACTGGGAA 343
QY 733 AGCTGAACCTTGACCAGACTCATTTAGAGACAGTAATTCAGACACCAAGGAAAGAAAT 792
DB 342 AAGCTAAAAAGTGATCAAGCTCATTTGGAGACGGTAATCCCCGCTTTGGACAAGCCTGTC 283
QY 793 CTAGTTTTAAATGGAGGCTACAGAGAAATGAAGGTACCTTAGAATCCATCAATGAGAAG 852
DB 282 ATGGTGGTTAATGGCGCTTATCGGGATCCGAGGCTCTGCTAAGGAAACTGGACGCGC 223
QY 853 ACTTTTTCAGTACTATCGTCAATTGAACCTGGCCCTTTAAAGGACGAGAGTTGAAGGA 912
DB 222 AGATATTCAAGTCAGCGTGGAATATTGACGGTCTCTCAAAGGCAGAAATTTGAGACAAC 163
QY 913 ATTCATATGACACATTTCTAAACT 938
DB 162 GTGCAATACGAAGATATATCTAAACT 137

RESULT 3
US-09-270-767-15661/c
; Sequence 15661, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15661
; LENGTH: 538
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-15661

Query Match 12.3%; Score 123.6; DB 3; Length 538;
Best Local Similarity 57.5%; Pred. No. 2.2e-26;
Matches 222; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

QY 553 ATCATGGAGATTGAAGAGGAAAGAAAGAACTGCCCGAACACAGACTACTGGCTACAGCCT 612
DB 522 ATCAAGCAGGAGAAAGCAAAAGGAGCGTGCACCGCAAAAGACTACTGGCTGCACAA 463
QY 613 GAAATATTGTGAAATATTAAACCAAGAACTGGGAGAGAAATATCATAGAAAAAGGCT 672

DB 462 GGTATCGTGTCAAAATTTATTTCCAAATCCATGGCGGAAAGTTCTTCAAAACAAAAAGCG 403
QY 673 ATGTGTAAGAAAGTAATTGACAAATATACAGCTGTTGTGAAGATGATGATCTTGAGAC 732
DB 402 GTTGTCTCGAGCTAATTGACAGATATCAGGCGCAAAATCAAGTTCTTGAGAGACTGGGAA 343
QY 733 AAGCTGAACCTTGACCAGACTCATTTAGAGACAGTAATTCAGACACCAAGGAAAGAAAT 792
DB 342 AAGCTAAAAAGTGATCAAGCTCATTTGGAGACGGTAATCCCCGCTTTGGACAAGCCTGTC 283
QY 793 CTAGTTTTAAATGGAGGCTACAGAGAAATGAAGGTACCTTAGAATCCATCAATGAGAAG 852
DB 282 ATGGTGGTTAATGGCGCTTATCGGGATCCGAGGCTCTGCTAAGGAAACTGGACGCGC 223
QY 853 ACTTTTTCAGTACTATCGTCAATTGAACCTGGCCCTTTAAAGGACGAGAGTTGAAGGA 912
DB 222 AGATATTCAAGTCAGCGTGGAATATTGACGGTCTCTCAAAGGCAGAAATTTGAGACAAC 163
QY 913 ATTCATATGACACATTTCTAAACT 938
DB 162 GTGCAATACGAAGATATATCTAAACT 137

RESULT 4
US-09-248-796A-5938
; Sequence 5938, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 5938
; LENGTH: 825
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-5938

Query Match 7.9%; Score 79; DB 3; Length 825;
Best Local Similarity 52.8%; Pred. No. 5e-13;
Matches 220; Conservative 0; Mismatches 190; Indels 7; Gaps 2;

QY 57 CATGGGGAAGTCGGATTTTCTTACTCCCAAGGCTATCCCAACAGGATCAAGTCCAAGG 116
DB 105 CATGCAAAAGGCAGAGTTTGGAAACAGCAAAATATCAATCTAAAAAGCTACGAGCGCAG 164
QY 117 GCTGCAGAAAGCTACGCTGGTATGCGATGTCAGAGAGCAGTCCGGGACGAGAAATGG 176
DB 165 ATTACAAAAATTTGAAGTTCTATTGCGCAATTTGTTCTTAAACAGTGTGAGATTTCAAT 224
QY 177 CTTTAAAGTGTCAATGTATGTCGGAATCTCATCAGAGACAACT-----ATTCTGGCTTC 230
DB 225 GTTTAAAAACCAATCTATCTTACCGCTGCACATAAAGAGGTTTCCGAAATACATGAATC 284
QY 231 AGAAAAATCCTCAGCAGTTTATGGAATTTATTTTCAGAGAAATTTCCGAATGATCTTCTAGA 290
DB 285 TGGTGATAGTCCAAACTAATAGAAACCTACTCTACAAAAATTTCCAAGATAAGTTTATCAA 344
QY 291 ACTTCTCAGGAGACGCTTTGGCACTAAAAGGGTCCACAAACATTTGCTCAACGAATA 350
DB 345 GTTGTCTCCGAATCAACCACTGGAACCTAAAATTTATCAATGCCAATAAGTTCTACCAAGAATA 404
QY 351 CATCAGCCACCGAGAGACATCCACATGAATGCCCACTCAGTGGGAAACTCTGACTGATTT 410
DB 405 TATCCGCGAGCGAGATCATATCCATATGAAATTCACCCAGATGGAGAGTCTAACTTCATT 464

US-10-071-411A-63
; Sequence 63, Application US/10071411A
; Patent No. 6797475
; GENERAL INFORMATION:
; APPLICANT: Joanne Meyer
; TITLE OF INVENTION: Detection of Polymorphisms in the Human
; FILE REFERENCE: MRI-021
; CURRENT APPLICATION NUMBER: US/10/071,411A
; CURRENT FILING DATE: 2002-02-07
; PRIOR FILING DATE: 2002-02-07
; PRIOR FILING DATE: 2001-02-08
; PRIOR APPLICATION NUMBER: 60/314,248
; PRIOR FILING DATE: 2001-08-21
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63
; LENGTH: 168174
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(168174)
; OTHER INFORMATION: n = A,T,C or G
US-10-071-411A-63

Query Match 4.2%; Score 42.2; DB 3; Length 168174;
Best Local Similarity 46.6%; Pred. No. 0.85;
Matches 169; Conservative 0; Mismatches 193; Indels 1; Gaps 1;
QY 498 CCAGAGCTCAACTCAGTCTTAAGAAAGAGAAAGAACTGCCGACAGACTACTGGCTACAGCTGAAAT 557
Db 146251 CAAGAGTGAACCTCTGCTCAAAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 146310
QY 558 GGAGATTGAAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 617
Db 146311 GAAGAAAGAGAGAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 146370
QY 618 TATTGTGAATTTATACCAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 677
Db 146371 AAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 146430
QY 678 TAAGGAAGTATTCAGAAATATACAGCTGTTGTGAAGATGATTCATTCAGAGCAAGCT 737
Db 146431 AAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 146490
QY 738 GAAGCTTGCAGAGCTCATTTAGAGACAGTAAATTCAGACCAAGAAAGAAAGAAAGAAAGTCTAGT 797
Db 146491 AAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 146549
QY 798 TTTAAATGGAGGCTACAGAGAAATGAAGGTACCCCTAGAAATCCATCAATGAGAGACTTT 857
Db 146550 GAAAGAAAGAAAGAAAG 146609
QY 858 TTC 860
Db 146610 TTC 146612

RESULT 11
US-10-071-411A-2
; Sequence 2, Application US/10071411A
; Patent No. 6797475
; GENERAL INFORMATION:
; APPLICANT: Joanne Meyer
; TITLE OF INVENTION: Detection of Polymorphisms in the Human
; FILE REFERENCE: MRI-021
; CURRENT APPLICATION NUMBER: US/10/071,411A
; CURRENT FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: 60/267,515

; PRIOR FILING DATE: 2001-02-08
; PRIOR APPLICATION NUMBER: 60/314,248
; PRIOR FILING DATE: 2001-08-21
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 168273
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(168273)
; OTHER INFORMATION: n = A,T,C or G
US-10-071-411A-2
Query Match 4.2%; Score 42.2; DB 3; Length 168273;
Best Local Similarity 46.8%; Pred. No. 0.85;
Matches 169; Conservative 0; Mismatches 193; Indels 1; Gaps 1;
QY 498 CCAGAGCTCAACTCAGTCTTAAGAAAGAGAAAGAAATCTGCATCGATGAAATCAT 557
Db 146350 CAAGAGTGAACCTCTGCTCAAAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 146409
QY 558 GGAGATTGAAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 617
Db 146410 GAAAGAAAGAGAGAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 146469
QY 618 TATTGTGAATTTATACCAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGCTATTGT 677
Db 146470 AAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 146529
QY 678 TAAGGAAGTATTCAGAAATATACAGCTGTTGTGAAGATGATTCATTCAGAGCAAGCT 737
Db 146530 AAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 146589
QY 738 GAAGCTTGCAGAGCTCATTTAGAGACAGTAAATTCAGACCAAGAAAGAAAGAAAGTCTAGT 797
Db 146590 AAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 146648
QY 798 TTTAAATGGAGGCTACAGAGAAATGAAGGTACCCCTAGAAATCCATCAATGAGAGACTTT 857
Db 146649 GAAAGAAAGAAAGAAAG 146708
QY 858 TTC 860
Db 146709 TTC 146711

RESULT 12
US-09-949-016-89871/c
; Sequence 89871, Application US/09949016
; Patent No. 8812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 89871
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-89871

Query Match 4.2%; Score 42; DB 3; Length 601;
Best Local Similarity 55.7%; Pred. No. 0.041;
Matches 78; Conservative 1; Mismatches 61; Indels 0; Gaps 0;

QY 840 CATCAATGAGAAGACTTTTTCAGCTACTATCGTCATTGAAACTGGGCCCTTTAAAGGACG 899
DB 385 CTTCACTGGAGCTGCTTCTTCAGTTACTCTATTCACAGCAAGTTGAACCTCTAAAGGAA 326

QY 900 CAGAGTTGAAGAAATCAATATGAGACATTTCTTAAACTTGGCTGAGTTGAAAATTTGT 959
DB 325 AAGAGTTTAAACATTACAGAAATTCMATTTCTTCTTAAAACTTCCAAAGATTAAACATGACT 266

QY 960 TAACAATACCTTTAAATCT 979
DB 265 CATCCAAAACGATTTTACCT 246

RESULT 13
US-09-949-016-162984/c
; Sequence 162984, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 162984
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-162984

Query Match 4.2%; Score 42; DB 3; Length 601;
Best Local Similarity 55.7%; Pred. No. 0.041;
Matches 78; Conservative 1; Mismatches 61; Indels 0; Gaps 0;

QY 840 CATCAATGAGAAGACTTTTTCAGCTACTATCGTCATTGAAACTGGGCCCTTTAAAGGACG 899
DB 385 CTTCACTGGAGCTGCTTCTTCAGTTACTCTATTCACAGCAAGTTGAACCTCTAAAGGAA 326

QY 900 CAGAGTTGAAGAAATCAATATGAGACATTTCTTAAACTTGGCTGAGTTGAAAATTTGT 959
DB 325 AAGAGTTTAAACATTACAGAAATTCMATTTCTTCTTAAAACTTCCAAAGATTAAACATGACT 266

QY 960 TAACAATACCTTTAAATCT 979
DB 265 CATCCAAAACGATTTTACCT 246

RESULT 14
US-09-433-699-3
; Sequence 3, Application US/09433699B
; Patent No. 6165786
; GENERAL INFORMATION:
; APPLICANT: Lex M. Cowsett
; APPLICANT: Lex M. Cowsett
; TITLE OF INVENTION: ANTISENSE MODULATION OF NUCLEOLIN EXPRESSION
; FILE REFERENCE: R1S-0109
; CURRENT APPLICATION NUMBER: US/09/433,699B
; CURRENT FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 3
; LENGTH: 2518

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (112)..(2235)
US-09-433-699-3

Query Match 4.2%; Score 41.6; DB 3; Length 2518;
Best Local Similarity 50.5%; Pred. No. 0.12;
Matches 101; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 591 AACGACTACTGGCTACAGCCTGAAATTTATTGTGAAAATTTATAACCAAGAAACTGGGAGA 650
DB 753 ACCAGCCAAAGAAAGAAAGCTGCAAAAGTTGTTCTGTGAAAGCCNAGAACGTGGCTGA 812

QY 651 GAAATATCATAGAAAAAGGCTATTGTTAAGAAAGTAATTTGACAAATATACAGCTGTTGT 710
DB 813 GGATGAAGATGAAGAAGAGGATGATGAGGACGAGGATGACGACGACGAGAGATGATGA 872

QY 711 GAAGATGATTGATTTCTGGAGACNAGCTGAAACTTGACCACTCATTTTAGACACAGTAAT 770
DB 873 AGATGATGATGATGAAGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 932

QY 771 TCCAGCACGAGAAAAAGAA 790
DB 933 AGAAGCACCTGGAAAAACGAA 952

RESULT 15
US-09-949-016-4169
; Sequence 4169, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4169
; LENGTH: 2527
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4169

Query Match 4.2%; Score 41.6; DB 3; Length 2527;
Best Local Similarity 50.5%; Pred. No. 0.12;
Matches 101; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 591 AACGACTACTGGCTACAGCCTGAAATTTATTGTGAAAATTTATAACCAAGAAACTGGGAGA 650
DB 753 ACCAGCCAAAGAAAGAAAGCTGCAAAAGTTGTTCTGTGAAAGCCNAGAACGTGGCTGA 812

QY 651 GAAATATCATAGAAAAAGGCTATTGTTAAGAAAGTAATTTGACAAATATACAGCTGTTGT 710
DB 813 GGATGAAGATGAAGAAGAGGATGATGAGGACGAGGATGACGACGACGAGAGATGATGA 872

QY 711 GAAGATGATTGATTTCTGGAGACNAGCTGAAACTTGACCACTCATTTTAGACACAGTAAT 770
DB 873 AGATGATGATGATGAAGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 932

QY 771 TCCAGCACGAGAAAAAGAA 790
DB 933 AGAAGCACCTGGAAAAACGAA 952

Search completed: November 27, 2005, 01:10:41
Job time : 210.505 secs

This Page Blank (usp1c)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:00:41 ; Search time 1155.14 Seconds
(without alignments)
7173.086 Million cell updates/sec

Title: US-09-555-529-3

Perfect score: 1002

Sequence: 1 tgattcgactggtaccg.....agcatcaattggtgttcgc 1002

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.Main.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	371.2	37.0	461	7	US-10-242-535A-29177, A
2	371.2	37.0	461	7	US-10-085-783A-29177, A
3	267.2	26.7	591	5	US-10-106-698-1187, Ap
4	177.6	17.7	538	7	US-10-437-963-89433, A
5	176	17.6	1713	8	US-10-739-930-2955, Ap
6	174.4	17.4	2046	8	US-10-425-115-157125, A
7	171	17.1	300	9	US-10-779-543-7646, Ap
8	170.8	17.0	1614	7	US-10-425-114-2975, Ap
9	170.8	17.0	1713	7	US-10-425-114-22614, A
10	170	17.0	777	7	US-10-424-599-25484, A
11	163.4	16.3	1241	10	US-11-097-143-41039, A
12	163.4	16.3	3313	10	US-11-097-143-41038, A
13	163.4	16.3	4394	10	US-11-097-143-41017, A
14	163.4	16.3	4582	10	US-11-097-143-24532, A
15	156.8	15.6	584	7	US-10-767-701-25587, A
16	126.4	12.6	549	3	US-09-991-936-1424, Ap
17	126.4	12.6	549	9	US-10-978-245-1424, Ap
18	106.8	10.7	739	7	US-10-424-599-53324, A
19	102.4	10.2	516	4	US-09-925-065A-549989, A
20	100.2	10.0	270	3	US-09-294-093B-3089, Ap
21	72.6	7.2	1608	4	US-09-925-065A-707309, Ap
22	71.2	7.1	616	7	US-10-424-599-53326, A
23	67	6.7	588	4	US-09-925-065A-20834, A

ALIGNMENTS

RESULT 1

US-10-242-535A-29177
; Sequence 29177, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liaw, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 29177
; LENGTH: 461
; TYPE: DNA
; ORGANISM: Human
US-10-242-535A-29177

Query Match 37.0%; Score 371.2; DB 7; Length 461;

Best Local Similarity 91.2%; Pred. No. 1.4e-88;
Matches 405; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY 51 GGTCGCCATGGGGAAGTCGGAATTTCTTACTCCAAAGCTATCCCAAGCATCAAGTC 110
Db 1 GGTCGCCATGGGGAAGTCGGAATTTCTTACTCCCAAGCTATCCCAAGCATCAAGTC 59
QY 111 CAAGGGGCTGAGAAGCTACGCTGGTATGCGAGATGCGCAGAGCATGCCGGAGCA 170
Db 60 CAAGGGGCTGAGAAGCTACGCTGGTATGCGAGATGCGCAGAGCATGCCGGAGCA 119
QY 171 GAATGGCTTTAAAGTGTCAATGTATGTCCGAATCTCATCAGAGACAATATTGCTGGCTTC 230
Db 120 GAATGGCTTTAAAGTGTCAATGTATGTCCGAATCTCATCAGAGACAATATTGCTGGCTTC 179
QY 231 AGAAATCCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTCGGAATGACTTTCTAGA 290
Db 180 AGAAATCCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTCGGAATGACTTTCTAGA 239

QY 291 ACTTCTCAGGAGCGCTTTGGGCTCTAAAGGGTCCACAACAATTTGTCTTACAAACAATA 350
DB 240 ACTTCTCAGGAGCGCTTTGGGCTCTAAAGGGTCCACAACAATTTGTCTTACAAACAATA 299
QY 351 CATCAGCCACCGAGAGACATCCACATGAATGCCACTCAGTGGGAACCTCTGACTGATTT 410
DB 300 CATCAGCCACCGAGAGACATCCACATGAATGCCACTCAGTGGGAACCTCTGACTGATTT 359
QY 411 TACTAAGTGGCTGGGCGAGAGAGCGCTTTGTGCAAAAGTGCACCTGAAGACGATAGGAATTC 470
DB 360 TACTAAGTGGCTGGGCGAGAGAGCGCTTTGTGCAAAAGTGCACCTGAAGACGATAGGAATTC 419
QY 471 AGCATCAGTGAACGAAAGAATC 494
DB 420 ATATTCACTACATAGACAGGGACC 443

RESULT 2

US-10-085-783A-29177
; Sequence 29177, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29177
; LENGTH: 461
; TYPE: DNA
; ORGANISM: Human
US-10-085-783A-29177

Query Match 37.0%; Score 371.2; DB 7; Length 461;
Best Local Similarity 91.2%; Pred. No. 1.4e-88;
Matches 405; Conservative 0; Mismatches 38; Indels 1; Gaps 1;
QY 51 GCTCGCATGGGGAAGTCGGATTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTC 110
DB 1 GGTGCGCATGGGGAAGTCGGATTTCTTACTCCCAA-GCTATGCGCAA CAGGATCAAGTC 59
QY 111 CAAGGGGCTGCAGAGCTACGCTGGTATTGCCAGATGTGCCAGACAGTGC CGGGACGA 170
DB 60 CAAGGGGCTGCAGAGCTACGCTGGTATTGCCAGATGTGCCAGACAGTGC CGGGACGA 119
QY 171 GAATGCGCTTAAGTGTCATTTGATGTCGGAATCTCATCAGACAACACTATTGCTGGCTTC 230
DB 120 GAATGCGCTTAAGTGTCATTTGATGTCGGAATCTCATCAGACAACACTATTGCTGGCTTC 179
QY 231 AGAAATCTCAGCAGTTTATGATTTATTTTTCAGAGGAATTCGGAATGACTTTCTAGA 290
DB 180 AGAAATCTCAGCAGTTTATGATTTATTTTTCAGAGGAATTCGGAATGACTTTCTAGA 239
QY 291 ACTTCTCAGGAGCGCTTTGGGCTCTAAAGGGTCCACAACAATTTGTCTTACAAACAATA 350
DB 240 ACTTCTCAGGAGCGCTTTGGGCTCTAAAGGGTCCACAACAATTTGTCTTACAAACAATA 299
QY 351 CATCAGCCACCGAGAGACATCCACATGAATGCCACTCAGTGGGAACCTCTGACTGATTT 410
DB 300 CATCAGCCACCGAGAGACATCCACATGAATGCCACTCAGTGGGAACCTCTGACTGATTT 359
QY 411 TACTAAGTGGCTGGGCGAGAGAGCGCTTTGTGCAAAAGTGCACCTGAAGACGATAGGAATTC 470

DB 360 TACTAAGTGGCTGGGCGAGAGAGCGCTTTGCAAAAGTGCAGACGACCAAAAGGCTGTT 419
QY 471 AGCATCAGTGAACGAAAGAATC 494
DB 420 ATATTCACTACATAGACAGGGACC 443

RESULT 3

US-10-106-698-1187
; Sequence 1187, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 1187
; LENGTH: 591
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (587)..(587)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-1187

Query Match 26.7%; Score 267.2; DB 5; Length 591;
Best Local Similarity 95.3%; Pred. No. 1.1e-60;
Matches 286; Conservative 0; Mismatches 13; Indels 1; Gaps 1;

QY 703 GCTGTTGTGAAGATGATTCTCGAGACAAGCTTGAACACTTGACACAGCTCATTTAGAG 762
DB 28 GCTGCAGGAATTCGCGACGAGCTCGAGACAAGCTTGAACACTTGACACAGCTCATTTAGAG 87
QY 763 ACAGTAATTCGACACCGAGGAAAGAAATTCAGTTTTTAAATGGAGGCTACAGAGAAAT 822
DB 88 ACAGTAATTCGACACCGAGGAAAGAAATTCAGTTTTTAAATGGAGGCTACAGAGAAAT 147
QY 823 GAAGGTACCTTAGAATCCATCAATGAGAGACTTTTTCAGCTACTATCGTCAATGAAACT 882
DB 148 GAAGGTACCTTAGAATCCATCAATGAGAGACTTTTTCAGCTACTATCGTCAATGAAACT 207
QY 883 GGCCCTTTAAAGGACGAGAGTTGAAGGAATTCATATGAAGACATTTCTAAACTTGCC 942
DB 208 GGCCCTTTAAAGGACGAGAGTTGAAGGAATTCATATGAAGACATTTCTAAACTTGCC 267
QY 943 TGAGTTTGAAATTTGTTTAAACAATACCTTTTAAATCTTTAAAGCATCAAATTCGTTTCGC 1002
DB 268 TGAGTTTGAAATTTGTTTAAACAATACCTTTTAAATCTTTAAAGCATCAAATTCGTTTCGC 326

RESULT 4

US-10-437-963-89433
; Sequence 89433, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 89433
; LENGTH: 538
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_88190C.1
US-10-437-963-89433

Query Match 17.7%; Score 177.6; DB 7; Length 538;
Best Local Similarity 63.7%; Pred. No. 9.7e-37;
Matches 270; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 46 GCGGTGTGCGCCATGGGAGTCCGATTTCTTACTCCCAAGGCTATCGCAACAGGATC 105
DB 112 GTCCGGAAGAGATGGGAAGCAGAGTTCTCGACGCGGAGGCGATCGGAACAGATC 171

QY 106 AAGTCCAAGGGGCTGCAGAAAGTACGCTGGTATTGCGAGATGCGCAGAGCAGTCCCGG 165
DB 172 AAGCGAAGGGCTGCAGAAAGTCCGCTGCTACTGCGCAGATGCGCAGAGCAGTCCCGC 231

QY 166 GACGAGAATGGCTTAAAGTGTCAATGTATGTCGGAATCTCATCAGAGAACATTTGCTG 225
DB 232 GACGAGAATGGCTTCAAGTGCCTACATGTCGAGTGCACAGCGCCAGATGCGAGGTG 291

QY 226 GCTTCAGAAATCTCAGCAGTTATGATTTATTTTTCAGAGGAATTCGGAATGACTTT 285
DB 292 TTCGGCAGCGCCCGCAGCGTCTGAGGGCTTCTCCGAGGAGTTCTTCGAGCGCTTC 351

QY 286 CTAGAACTTCTCAGGAGACGCTTGGCACTAAAAGGCTCCACAAACATTTGTCTACAAC 345
DB 352 CTACCTTGTCTCGCCGCGCCACCGACACTCCCGCATCGCCGCGCTCGTCTACAAC 411

QY 346 GAATACATCAGCCACCGAGAGCAGATCCATGAATGCCACTCAGTGGGAACTCTGACT 405
DB 412 GAGTTTATCGCCAGCCACCGTCCATGAATCTCCACGGCTGGGCGCAGCTCACC 471

QY 406 GATTTTACTAAGTGGCTGGGAGAGGCTTGGCAAAAGTGCATGAGAGCATAGGA 465
DB 472 GAGTTTCTCAAGTTCTCTCGGCGCGAGGCGCATGCAAGGTGAGGACACTCCCAAAGG 531

QY 466 AGTT 469
DB 532 TGGT 535

RESULT 5
US-10-739-930-2955
; Sequence 2955, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 2955
; LENGTH: 1713
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: ZEAMA-23APR03-CLUSTER7583_2
US-10-739-930-2955

Query Match 17.6%; Score 176; DB 8; Length 1713;
Best Local Similarity 63.9%; Pred. No. 4.9e-36;

Matches 266; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

QY 54 CGCCATGGGGAAGTCCGATTTTCTTACTCCCAAGGCTATCGCAACAGGATCAAGTCCAA 113
DB 171 CCGATGGGGAAGCAGAGTTCTCGACGCGGAAGGGATCGGAACCGGATCAAGGCGAA 230

QY 114 GGGGCTGCAGAAAGTACGCTGGTATTTGCGAGATGCGCAGAGCAGTGGCGGAGCAGAA 173
DB 231 GGGGCTGCAGAAAGTACGCTGGTATTTGCGAGATGCGCAGAGCAGTGGCGGAGCAGAA 290

QY 174 TGGCTTTAAGTGTCTATTGTATGTCGGAATCTCATCAGAGACAACTATTCTGCTTCAGA 233
DB 291 CCGGTTCAAGTGCCTACTGATGCGAGTCCGACCGGCGAGATGCAAGTGTTCGCGCAT 350

QY 234 AAATCTCTCAGCAGTTTATGATTTATTTTTCAGAGGAATTCGGAATGACTTTCTAGAACT 293
DB 351 GGGGCGCGGACCGGCTCGTCCGAGGGCTTCTCCGAGGAATTCCTCGAGTCTTCTCTCCCT 410

QY 294 TCTCAGGAGACGCTTTTGGCACTAAAAGGTCACAAACATTTGTCTACAAAGAAATACAT 353
DB 411 CATCCGCGCGCGCACCGCCACTCCCGCTCGCGCCACCGTTGTCTACAAAGATACAT 470

QY 354 CAGCCACCGAGAGCAGATCCATGAATGCCACTCAGTGGGAACTCTGACTGATTTTAC 413
DB 471 CGCCGACCGGCGCACCGTCCATGAACTCTACGCGTGGGCCACGCTCACCAGTTCGT 530

QY 414 TAAAGTGGCTGGCAGAGAGGCTTGTGCAAAAGTGCATGAAGCAGATAGGAAGTT 469
DB 531 CAAAGCTCTGGGGCGGGAAGGGTACTGCAAGGTTGAGGACACGCCCAAGGGGTGGT 586

RESULT 6
US-10-425-115-157125
; Sequence 157125, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 157125
; LENGTH: 2046
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_74879C.1
US-10-425-115-157125

Query Match 17.4%; Score 174.4; DB 8; Length 2046;
Best Local Similarity 63.7%; Pred. No. 1.4e-35;
Matches 265; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

QY 54 CGCCATGGGGAAGTCCGATTTTCTTACTCCCAAGGCTATCGCAACAGGATCAAGTCCAA 113
DB 294 CCGATGGGGAAGCAGAGTTCTCGACGCGGAAGGGATCGGAATCGGATCAAGGCGAA 353

QY 114 GGGGCTGCAGAAAGTACGCTGGTATTTGCGAGATGCGCAGAGCAGTGGCGGAGCAGAA 173
DB 231 GGGGCTGCAGAAAGTACGCTGGTATTTGCGAGATGCGCAGAGCAGTGGCGGAGCAGAA 413

QY 174 TGGCTTTAAGTGTCTATTGTATGTCGGAATCTCATCAGAGACAACTATTGTGCTTCAGA 233
DB 414 CCGGTTCAAGTGCCTACTGATGCGAGTCCGACCGGCGAGATGCAAGTGTTCGCGCAT 473

QY 234 AAATCTCTCAGCAGTTTATGATTTATTTTTCAGAGGAATTCGGAATGACTTTCTAGAACT 293
DB 474 GGGGCGCGGCGCGCTCGTCCAGGGCTTCTCCGAGGAATTCCTCGAGTCTTCTCTCCCT 533

QY 294 TCTCAGGACGCTTTGGCACTAAAGGTCACAAACAATGTCTACAAAGATACAT 353
Db |||||
534 CATCGCCGCGCACCGCACTCCCGCGTCGCGCGTGTCTACAAAGATACAT 593
QY 354 CAGCCACCGAGACACATCCACATGAATGCCACTCAGTGGGAACTCTCAGCTGATTTTAC 413
Db |||||
594 CGCGACCGGACACGCTCCACATGAATCTACGCGGTGGGCCACGCTCACCGAGTTCTGT 653
QY 414 TAAAGTGGCTGGGACAGAGGCTTGTGCAAAAGTGCACCTGAAGACGATAGGAAGTT 469
Db |||||
654 CAAGCTCTGGGCGCGGAAGGTTACTGCAAGGTTGAGGACACGCGCCCAAGGGTGTGT 709

RESULT 7

US-10-779-543-7646
; Sequence 7646, Application US/10779543
; Publication No. US20050227917A1
; GENERAL INFORMATION:
; APPLICANT: Williams et al
; TITLE OF INVENTION: GENE PRODUCTS DIFFERENTIALLY EXPRESSED
; FILE REFERENCE: 2300-21302
; CURRENT APPLICATION NUMBER: US/10/779,543
; PRIORITY FILING DATE: 2004-02-12
; PRIORITY FILING DATE: 2002-02-15
; PRIORITY FILING DATE: 09/217,471
; PRIORITY FILING DATE: 1998-12-21
; PRIORITY FILING DATE: 60/068,755
; PRIORITY FILING DATE: 1997-12-23
; PRIORITY FILING DATE: 60/080,664
; PRIORITY FILING DATE: 1998-04-03
; PRIORITY FILING DATE: 60/105,234
; PRIORITY FILING DATE: 1998-10-21
; PRIORITY FILING DATE: 09/297,648
; PRIORITY FILING DATE: 2000-04-10
; PRIORITY FILING DATE: 1999-01-28
; PRIORITY FILING DATE: 60/072,910
; PRIORITY FILING DATE: 1998-01-28
; PRIORITY FILING DATE: 60/075,954
; PRIORITY FILING DATE: 1998-02-24
; PRIORITY FILING DATE: 60/080,114
; PRIORITY FILING DATE: 1998-03-31
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 23767
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7646
; LENGTH: 300
; TYPE: DNA
; ORGANISM: Homo sapiens

US-10-779-543-7646

Query Match 17.1%; Score 171; DB 9; Length 300;
Best Local Similarity 100.0%; Pred. No. 4.1e-35;
Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 445 AGTGCACTGAAGACGATAGGAAGTTTCAGCATCAGTGAAGAGAAAGAAATCTTCCCGAGAGC 504
Db |||||
130 AGTGCACTGAAGACGATAGGAAGTTTCAGCATCAGTGAAGAGAAAGAAATCTTCCCGAGAGC 189
QY 505 TCAACTCAGTCTAAAGAGAAAGAGAAAGAAATCTTGCATCGATGAATCATGGAGATT 564
Db |||||
190 TCAACTCAGTCTAAAGAGAAAGAGAAAGAAATCTTGCATCGATGAATCATGGAGATT 249
QY 565 GAAGAGAGAAAGAGAAAGAGAAAGAGAAATCTTGCATCGATGAATCATGGAGATT 615
Db |||||
250 GAAGAGAGAAAGAGAAAGAGAAAGAGAAATCTTGCATCGATGAATCATGGAGATT 300

RESULT 8

US-10-425-114-2975

; Sequence 2975, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 2975
; LENGTH: 1614
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700238628_FLI
US-10-425-114-2975

Query Match 17.0%; Score 170.8; DB 7; Length 1614;
Best Local Similarity 64.9%; Pred. No. 1.2e-34;
Matches 253; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 54 CGCCATGGGGAAGTCGGATTTTCTTACTCCCAAGGCTATCGCAAACAGGATCAAGTCCAA 113

Db |||||

76 CACGATGGGGAAGCACGAGTTCTTGACGCCGAAGCGGATCGCAAACCGGATCAAGGCGAA 135

QY 114 GGGGCTGCAGAAAGCTACCTGGTATTGCCAGATGCCAGAAAGAGTCCGGGACGAGAA 173

Db |||||

136 GGGGCTGCAGAAAGCTACCTGGTATTGCCAGATGCCAGAAAGAGTCCGGGACGAGAA 195

QY 174 TGGCTTTAAGTGTCAATGTATGTCGAATCTCATCAGAGACAACTATTGCTGGCTTCAG 233

Db |||||

196 CGGTTTCAAGTGTCCACTGCAATGTCGAGTCCGACGAGGAGATGCGAGTTCGGCAT 255

QY 234 AAATCCTCAGCAGTTTATGGATTATTTTCAGAGAAATTCGAAATGACTTTCTAGAACT 293

Db |||||

256 GGGCCCGACCGCTCGTCGAGGGCTTCTCGAGGAGTTCCTCGAGTCTCTCTCTCCCT 315

QY 294 TCTCAGGAGAGCTTTGGCACTAAAGGTCACAAACATTTGTCTACAGGATACAT 353

Db |||||

316 CATCGCGCGCGCACCGCCACTCCCGGTCGCGCCACCGTCTGCTCTACAGGATACAT 375

QY 354 CAGCCACCGAGAGACATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTTTAC 413

Db |||||

376 CGCCGACCGGACCACTCCACATGAACTCCAGCGGTGGGCCACGCTCACCGAGTTCTGT 435

QY 414 TAAAGTGGCTGGGACAGAGAGGCTTGTGCAA 443

Db |||||

436 CAAAGTTCTGGGGCGGAGGGGTACTGTAA 465

RESULT 9

US-10-425-114-22614
; Sequence 22614, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128


```
; SEQ ID NO 22614
; LENGTH: 1713
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3591-011-E3_FLI
US-10-425-114-22614

Query Match      17.0%; Score 170.8; DB 7; Length 1713;
Best Local Similarity 64.9%; Pred. No. 1.2e-34;
Matches 253; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

Qy 54 CGCCATCGGGAAGTCGGATTTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCAA 113
Db 235 CACGATGGGAGACACAGATTCCTGACCGCGAAGCGCATCGCAACCGGATCAAGGCGAA 294
Qy 114 GGGCTCAGAGAGTACGCTGGTATTGCCAGATGTCGCAAGCAGTGCCTGGGACGAGAA 173
Db 295 GGGCTCAGAGAGTGCCTGGTGGTACTGTCAGATGTCGCAAGCAGTGCCTGGGACGAGAA 354
Qy 174 TGGCTTTAAGTGTCTATTGTATGTCGGAATCTCATCAGACCAACTATTGCTGGCTTCGA 233
Db 355 CGGCTTCAAGTGCCTGTCGAGTCCGAGTCCGACGAGGAGATGCGAGTTCGGCAT 414
Qy 234 AATTCCTCAGCAGTTCATGATTTATTTTTCAGAGAAATTCGAAATGACTTCTAGAACT 293
Db 415 GGGCCCGGACCGCTGCTCGAGGCTTCTCGAGGAGTTCCTCGAGTCTTCTCTCCCT 474
Qy 294 TCTCAGGAGAGCGTTTGGCACTAAAGGCTCCACAAACAACTTGTCTACACGAATACAT 353
Db 475 CATCGCGCGCGCACCGCCACTCCCGCTCGCGCCACCGCTGCTCTACAAACGAGTACAT 534
Qy 354 CAGCCACCGAGAGACATCCATGAAATGCGCACTCAGTGGGAAATCTGACTGATTTTAC 413
Db 535 CGCGACCGGACCAACGCTCCACATGAACTCCACCGTGGGCCACGCTCACCGAGTTTCT 594
Qy 414 TAACTGCTGGGACAGAGCGCTTGTGCAA 443
Db 595 CAAGTTCCTGGGCGCGAGGGGTACTGTAA 624

RESULT 10
US-10-424-599-25484
; Sequence 25484, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424.599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 25484
; LENGTH: 777
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_123013C.1
US-10-424-599-25484

Query Match      17.0%; Score 170; DB 7; Length 777;
Best Local Similarity 57.5%; Pred. No. 1.3e-34;
Matches 305; Conservative 0; Mismatches 225; Indels 0; Gaps 0;

Qy 53 TCGCATGGGGAAGTCGGATTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCA 112
Db 135 TCGTATGGGCAAAAATGAGTTTCTCACCTAAAGCAATTCGCAATCGAATCAAGCAAA 194
Qy 113 AGGGCTCCCAAGTACGCTGGTATTGCCAGATGTGCCAAGCAGTGCCTGGGACGAGA 172
```

```
Db 195 AAGGATTGCAGAAGCTTCGGTGTATTTGCCAGATGTGTCCAGAGCAGTGCAGAGATGAGA 254
Qy 173 ATGGCTTTAAGTGTCAATTTATGTCGGAATCTCATCAGAGACAACTATTGCTGGCTTCAG 232
Db 255 ATGGGTTTAAATGCCAATTCATGAGTGAAGGCGCCAGCGTCAATGCGAGATTTTGGAC 314
Qy 233 AAAATCCTCAGCAGTTCATGATTTATTTTTCAGAGGAATTCGAAATGACTTTCTAGAAC 292
Db 315 AAAACCCACACCGGATAGTTGAGGGCTATTCCGAAGAGTTTTCAGAGTACTTTTCTGGAGC 374
Qy 293 TTCTCAGGAGACGCTTTGGCACTAAAGGGTTCACAAACATTTGCTTACAAAGATACA 352
Db 375 ACATGAAGGCGCAGTCCCGGATTCAGCCGCTGTGGCAGCCACTGTAGTTTATAACGAATACA 434
Qy 353 TCAGCCACCGAGAGACATCCATCAATGCACTCAGTGGGAACTCTGACTGATTTTA 412
Db 435 TAAATGACAGACACCACTCATATGAATCTTACTCAGTGGGCTACGCTTACTGAGTTTG 494
Qy 413 CTAAGTGGCTGGGCGACAGAGGCTTGTGCAAAAGTGCACCTGMAAGACGATAGAGGATTCAG 472
Db 495 TTAAGTACTTGGTTCGAACTGGCAATGTAAAGTTGAGGAAACACCAAGGGATGGTTCA 554
Qy 473 CATCAGTGAACGAAAGAAATCTTCCAGAGCTCAACTCAGTCTTAAAGAAAAGAGAA 532
Db 555 TTACATATATAGATAGAGATTCAGAAACCCCTTTTCAAGGAGAGGATGAAGATAAGAGAA 614
Qy 533 AGAAATCTGCACCTGGATGAATCATCGAGATTCAGAGGAGGAGGAGGAGGAGGAGGAG 582
Db 615 TCAAGGCAGATATGGTAGATGAAGAAAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 664

RESULT 11
US-11-097-143-41039
; Sequence 41039, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41039
; LENGTH: 1241
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-41039

Query Match      16.3%; Score 163.4; DB 10; Length 1241;
Best Local Similarity 64.9%; Pred. No. 9.7e-33;
Matches 242; Conservative 0; Mismatches 131; Indels 0; Gaps 0;
```

```
QY 56 CCATGGGAAGTCGGATTTTCTTACTCCCAAGGCTATGCCCAACAGGATCCCAAGTCCAAGG 115
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 67 CGATGGGTCCGCGGAGGTAGGTACGCCCAAGTACTCTGCCCAACAGATGAATCGAAGG 126
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 116 GGTCTGAGAAAGCTACGCTGGTATTTGTCAGATGTGCCAGAAGCAGTCCCGGGACGAGAATG 175
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 127 GTCTGAGAAAGCTCGCTGGTACTGCTCCAGATGTGCCAGAAGCAGTCCCGCGATGAAACG 186
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 176 GCTTTAAGTGTCAATGTATGTCCGAATCTCATCAGAGAACTCCGAAATGACTTTCTAGAACTTC 235
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 187 GCTTCAAGTGCACACAGTCCGAGTCCACACGAGCGCAGTTGCTCTCTTTGCGGACA 246
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 236 ATCCTCAGCAGTTTATGGATTTATTTTTCAGAGAAATTCGAAATGACTTTCTAGAACTTC 295
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 247 ATCCTGGCAAAATTCCTGSCACAGTTTCAGCAAGAGTTTCTCCGACGGCTACATGGAGTTGC 306
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 296 TCAGGAGAGCGCTTTGGCACTTAAAGGGTCCACAAACATTTGTCTACAACGAATACATCA 355
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 307 TGGCGCGCGGTTCCGACAGAGCAACAGCGCCCAACGAATCTTACAGGAGTACATTTG 366
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 356 GGCACCGAGAGACATCCATGAATGCGCACTCAGTGGGAAACTCTGACTGATTTTACTA 415
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 367 CCACAAGGAGACATCCATGAAGCCACCGCATGGTGTCAACCTGTCCGACTACGTGA 426
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 416 AGTGGCTGGCGAG 428
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 427 AGTGGCTGGGCGG 439
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
RESULT 12
US-11-097-143-41038
; Sequence 41038, Application US/11097143
; Publication No. US2005020858A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; PRIOR FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41038
; LENGTH: 3313
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-41038
```

```
Query Match 16.3%; Score 163.4; DB 10; Length 3313;
Best Local Similarity 64.9%; Pred. No. 1.7e-32;
Matches 242; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 56 CCATGGGAAGTCGGATTTTCTTACTCCCAAGGCTATGCCCAACAGGATCCCAAGTCCAAGG 115
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1067 CGATGGGTCCGCGGAGGTAGGTACGCCCAAGTACCTCGCCCAACAGATGAATCGAAGG 1126
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
QY 116 GGCTGCAGAACTACGCTGGTATTGCCAGATGTGCCAAGCAGTCCCGGACGAGAATG 175
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1187 GTCTGAGAAAGTCGCTGGTACTGCCAGATGTGGGAAGCAGTCCCGCATGAAAACG 1186
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 176 GCTTTAAGTGTCAATGTATGTCCGAATCTCATCAGAGAACTCCGAAATGACTTTCTAGAACTTC 235
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1187 GCTTCAAGTGCACACAGATGAGCGAGTCCACACGCGCCAGTTGCTCTCTTTGCGGACA 1246
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 236 ATCCTCAGCAGTTTATGGATTTATTTTTCAGAGGAATTCGGAATGACTTTCTAGAACTTC 295
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1247 ATCCTGGCAAAATTCCTGCACACGCTTCAGCAAAAGAGTTTCTCCGACGGCTACATGGAGTTGC 1306
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 296 TCAGGAGAGCGCTTTGGCACTTAAAGGGTCCACAAACATTTGTCTACAACGAATACATCA 355
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1307 TGGCGCGCGGTTCCGCAAGAGGAAACAGCGCCCAACGAATCTTACAGGAGTACATTTG 1366
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 356 GGCACCGAGAGACATCCATGAATGCACTCAGTGGGAAACTCTGACTGATTTTACTA 415
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1367 CCCACAAGGAGACATCCACATGAACGCCACCCGATGGCTCACCCCTGTCCGACTACGTGA 1426
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 416 AGTGGCTGGCGAG 428
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1427 AGTGGCTGGGCGG 1439
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
RESULT 13
US-11-097-143-41017
; Sequence 41017, Application US/11097143
; Publication No. US2005020858A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41017
; LENGTH: 4394
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-41017
```

```
Query Match 16.3%; Score 163.4; DB 10; Length 4394;
Best Local Similarity 64.9%; Pred. No. 1.9e-32;
Matches 242; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 56 CCATGGGAAGTCGGATTTTCTTACTCCCAAGGCTATGCCCAACAGGATCCCAAGTCCAAGG 115
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3756 CGATGGGTCCGCGCGGAGGTAGGTACGCCCAAGTACCTCGCCCAACAGATGAATCGAAGG 3815
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 116 GGCTGCAGAACTACGCTGGTATTGCCAGATGTGCCAAGCAGTCCCGGACGAGAATG 175
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Db 3816 GTCTGCAGAGCTGGCTGGTACTGCCAGATGTCGAGAGCAGTGGCGGATGAAACG 3875
Qy 176 GCTTTAAGTGTATGATGATCGAATCTCATCAGAGAACATTTGCTGGCTTCAGAAA 235
Db 3876 GCTTCAAGTGCACACGATGAGCGAGTCCACACGAGCCAGTTGCTCTCTTTTGGGACA 3935
Qy 236 ATCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTCGGAATGACTTTCTAGAACTTC 295
Db 3936 ATCTTGGCAATTCCTGCACAGCTTCAGCAAGAGTTCTCCGACGGCTACATGGAGTTGC 3995
Qy 296 TCAGGAGACCTTTGGCACTTAAAGGTCACCAACCAACATTTGTCTACAAAGATACATCA 355
Db 3996 TGGCGCGCGTTTCGGCACAGGACGACCAACCAACAGATCTTACAGGAGTACATTG 4055
Qy 356 GCACCCAGAGCAGATCCACATGAATGCCACTCAGTGGGAACTCTGACTGATTTACTA 415
Db 4056 CCACAGAGGAGCATCCATCATGAACCCAGCCGATGGCTCACCCTGTCTCCGACTACGTGA 4115
Qy 416 AGTGGCTGGGCG 428
Db 4116 AGTGGCTGGGCG 4128

RESULT 14
US-11-097-143-24532
; Sequence 24532, Application US/11097143
; Publication No. US2005020859A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24532
; LENGTH: 4582
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-24532

Query Match 16.3%; Score 163.4; DB 10; Length 4582;
Best Local Similarity 64.9%; Pred. No. 2e-32;
Matches 242; Conservative 0; Mismatches 131; Indels 0; Gaps 0;
Qy 56 CCATGGGGAAGTCGATTTCTTACTCCCAAGGCTATCCCAACAGGATCAAGTCCAAAG 115
Db 3944 COATGGGTGGCGGAGGTAGTACGCCCAAGTACCTGCCCAACAGATGAATCGAAG 4003
Qy 116 GGCTGCAGAGCTACGCTGGTATGTCAGAGATGTCAGAGCAGTGGCGGACAGGAATG 175
Db 4004 GTCTGCAGAGCTGCGCTGGTACTGCCAGATGTCGAGAGCAGTGGCGGATGAAACG 4063
Qy 176 GCTTTAAGTGTATGATGATCCGAACTCTCATCAGAGAACATTTGCTGGCTTCAGAAA 235

Db 4064 GCTTCAAGTGCACACGATGAGCGAGTCCACAGCGCCAGTTGCTCTCTTTGGCGACA 4123
Qy 236 ATCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTCGGAATGACTTTCTAGAACTTC 295
Db 4124 ATCTTGGCAATTCCTGCACAGCTTCAGCAAGAGTTCTCCGACGGCTACATGGAGTTGC 4183
Qy 296 TCAGGAGAGCTTTGGCACTTAAAGGTCACCAACCAACATTTGTCTACAAAGATACATCA 355
Db 4184 TGGCGCGCGTTTCGGCACAGGACGACCAACCAAGATCTTACAGGAGTACATTG 4243
Qy 356 GCACCCAGAGCAGATCCACATGAATGCCACTCAGTGGGAACTCTGACTGATTTACTA 415
Db 4244 CCACAGAGGAGCATCCACATGAACCCAGCCGATGGCTCACCCTGTCTCCGACTACGTGA 4303
Qy 416 AGTGGCTGGGCG 428
Db 4304 AGTGGCTGGGCG 4316

RESULT 15
US-10-767-701-25587
; Sequence 25587, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 25587
; LENGTH: 584
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 30963818
US-10-767-701-25587

Query Match 15.6%; Score 156.8; DB 7; Length 584;
Best Local Similarity 64.7%; Pred. No. 3.8e-31;
Matches 233; Conservative 0; Mismatches 127; Indels 0; Gaps 0;
Qy 58 ATGGGGAAGTCGAGTTTCTTACTCCCAAGGCTATCCCAACAGGATCAAGTCCAAAGGG 117
Db 225 ATGGGGAAGCAGAGTTCTTCAAGCCGAGCGATCCGAAACCGGATCAAGGCAAGGGG 284
Qy 118 CTCGAGAGCTACGCTGGTATGTCAGATGTCGAGAGCAGTCCCGGAGCAGAAATGGC 177
Db 285 CTGCAAAAGCTGGCGTGGTACTGTCAGATGTCGAGAGCAGTCCCGGAGCAGAAAGGG 344
Qy 178 TTTAAGTGTATGATGTCGAGATCTCATCAGAGCAACTATTGCTGGCTTCAGAAAT 237
Db 345 TTCAAGTGCCACTGCTGCGAGTCCACAGCGGAGATGCGAGTGTTCGGCATGCGC 404
Qy 238 CCTCAGCAGTTTATGGAATTTTTCAGAGAAATTCGAAATGACTTTCTAGAACTTCTC 297
Db 405 CCCGACCGGTCGTCGAGGGGCTTCTCGAGAGTTCTCTCGAGTCTCTCTCTCCCTCATC 464
Qy 298 AGGAGAGCTTTGGCACTTAAAGGTCACCAACCAACTTGTCTACACGAATACATCAGC 357
Db 465 CGCCGCGGCAACCGCCACTCCCGGCTGCCCGCCACCGTCTGCTACACGAGTACATCGC 524
Qy 358 CACCGAGAGCAGATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTTTACTAAG 417
Db 525 GACCGTACACAGTCCACATGNACTCCACGGCTGGGCCACGCTCAGCGAGTTTCTGTAAG 584

Result No.	Score	Query Match	Length	DB	ID	Description
1	44	4.4	437	1	US-10-821-234-402	Sequence 402, App
2	40.4	4.0	1575	7	US-11-074-176-145	Sequence 145, App
3	39.4	3.9	3254	1	US-10-793-626-4202	Sequence 4202, App
4	35.4	3.5	2631	7	US-11-074-176-321	Sequence 321, App
5	35.4	3.5	2649	7	US-11-074-176-87	Sequence 87, Appl
C 6	35	3.5	3043	1	US-10-793-626-3726	Sequence 3726, Ap
	34	3.4	3218	1	US-10-793-626-4329	Sequence 4329, Ap
7	33.4	3.3	3939	1	US-10-793-626-3103	Sequence 3103, Ap
8	33.4	3.3	3148	1	US-10-793-626-3369	Sequence 3369, Ap
9	33.4	3.3	3148	1	US-10-793-626-3369	Sequence 3369, Ap
10	32.4	3.2	1167	7	US-11-074-176-323	Sequence 323, App
11	32.4	3.2	1218	7	US-11-074-176-91	Sequence 91, Appl
12	32.4	3.2	3043	1	US-10-793-626-4176	Sequence 4176, Ap
13	31.8	3.2	2078	1	US-10-793-626-4417	Sequence 4417, Ap
C 14	31.8	3.2	125594	1	US-10-658-986-5	Sequence 5, Appl
15	31.4	3.1	2361	1	US-10-467-962B-102	Sequence 102, App
16	31.4	3.1	340000	7	US-11-102-978-3	Sequence 3, Appl
C 17	31	3.1	3073	1	US-10-793-626-3581	Sequence 3581, Ap
18	31	3.1	3494	1	US-10-793-626-3650	Sequence 3650, Ap
C 19	31	3.1	3926	1	US-10-793-626-4300	Sequence 4300, Ap
C 20	31	3.1	4138	1	US-10-793-626-3995	Sequence 3995, Ap
C 21	30.8	3.1	2975	1	US-10-793-626-3345	Sequence 3345, Ap
C 22	30.8	3.1	4114	1	US-10-793-626-4184	Sequence 4184, Ap
C 23	30.8	3.1	41948	1	US-10-949-720-393	Sequence 393, App

US-10-793-626-4202									
Query Match									
Best Local Similarity 3.9%; Score 39.4; DB 1; Length 3254;									
Matches 115; Conservative 0; Mismatches 126; Indels 0; Gaps 0;									
US-11-074-176-145									
; Sequence 145, Application US/11074176									
; Publication No. US20050250135A1									
; GENERAL INFORMATION:									
; APPLICANT: Klaenhammer, Todd R.									
; APPLICANT: Russell, William M.									
; APPLICANT: Altermann, Eric									
; APPLICANT: McAuliffe, Olivia									
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding									
; TITLE OF INVENTION: Stress-Related Proteins and Uses Therefore									
; FILE REFERENCE: 5051-694									
; CURRENT APPLICATION NUMBER: US/11/074,176									
; PRIOR FILING DATE: 2005-03-07									
; PRIOR APPLICATION NUMBER: 60/551,161									
; PRIOR FILING DATE: 2004-03-08									
; NUMBER OF SEQ ID NOS: 381									
; SOFTWARE: FastSeq for Windows Version 4.0									
; SEQ ID NO 145									
; LENGTH: 1575									
; TYPE: DNA									
; ORGANISM: Lactobacillus acidophilus									
; FEATURE:									
; NAME/KEY: CDS									
; LOCATION: (1)...(1575)									
; FEATURE:									
; NAME/KEY: misc feature									
; LOCATION: (0)...(0)									
; OTHER INFORMATION: ORF 1524; Lisk - Putative histidine kinase/may be									
; OTHER INFORMATION: Involved in stress response									
US-11-074-176-145									
Query Match									
Best Local Similarity 4.0%; Score 40.4; DB 7; Length 1575;									
Matches 83; Conservative 0; Mismatches 71; Indels 0; Gaps 0;									
QY	613	GAAATTTATTGTGAAATTAATTAACCAAGAACTGGGAGAGAAATATCATAGAAAAGGCT	672						
DB	1003	GAAATGCTTGATTAACTAGAGCTGAACAAATGATGTTCAATATCCTTATGAAGTAACA	1062						
QY	673	ATTGTTAAGGAGTAATTGACAAATATACAGCTGTTGTGAAGATGATTGATCTTGAGAC	732						
DB	1063	AATGTTAAACGAACAGTAACAAACGAGTGTCTCTGATTGGCAATGTTTCATTTCAGACTTT	1122						
QY	733	AAGCTGAACTTGACCAGACTCATTTAGAGACAG	766						
DB	1123	AAGATTCAACTTGATGAAGATGATTTACCACCAG	1156						
US-10-793-626-4202									
; Sequence 4202, Application US/10793626									
; Publication No. US20050255478A1									
; GENERAL INFORMATION:									
; APPLICANT: KIMMERLY, WILLIAM JOHN									
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS									
; FILE REFERENCE: PU3480US									
; CURRENT APPLICATION NUMBER: US/10/793,626									
; PRIOR FILING DATE: 2004-03-04									
; PRIOR APPLICATION NUMBER: 60/164,258									
; PRIOR FILING DATE: 1999-11-09									
; NUMBER OF SEQ ID NOS: 4472									
; SOFTWARE: PatentIn Ver. 2.1									
; SEQ ID NO 4202									
; LENGTH: 3254									
; TYPE: DNA									
; ORGANISM: Artificial Sequence									
; FEATURE:									
; OTHER INFORMATION: Description of Artificial Sequence: synthetic									
; OTHER INFORMATION: nucleic acid sequence									
US-11-074-176-321									
Query Match									
Best Local Similarity 45.1%; Pred. No. 2.3; DB 7; Length 2631;									
Matches 132; Conservative 0; Mismatches 161; Indels 0; Gaps 0;									
QY	518	AAGAAAAGAAGAAAAGAAATCTGCACCTGGATGAATCATGGAGATTGAAGAGGAAAAAGA	577						
DB	413	AAGAGGTCAGCTTAAATAGAGAGCTCAAAAAGCAAGAAAGAAATATCATGAGCATTGA	472						
QY	578	AAAGAACTGCCCGAACACAGACTACTGGCTACGCCCTGAAATATTGTGAAAAATTATAACCA	637						
DB	473	AGCATCCTAAAAAAGAGCAGTCAAGAAAAAGATAATAAGAAAAACAGTAAAGAAAGTAATA	532						
QY	638	AGAACTGGGAGAGAAATATCATAGAAAAAGGCTATTGTTAAGGAAGTAAATTGACAAAT	697						

Db 533 AAAAAAGTAGACACAAAGTGGAAAAAGAAAGTAATCGGACCAAAAAATCTTAAAAACCAT 592
Qy 698 ATACAGCTGTTGTAAGATGATTGATCTCGAGACAGCTGAAACTTGACCACTCAT 757
Db 593 CACCAGCCCGTTTAAAGAAAAATCAACCTGCAGATAATAAGGAAAAAGTTTACTACTCCTA 652
Qy 758 TAGACAGAGTAATTCACGACACACAGGAAAAAGAAATTCCTAGTTTAAATGGAGGC 810
Db 653 GAGTAACATATCTTGAAGACCAACCAAGGAGAAAAACGTGGTAAGGCTCGTGGC 705

RESULT 5

US-11-074-176-87
; Sequence 87, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
; APPLICANT: Klaenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Perill, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; FILE REFERENCE: Stress-Related Proteins and Uses Therefore
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074,176
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: 60/551,161
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 87
; LENGTH: 2649
; TYPE: DNA
; ORGANISM: Lactobacillus acidophilus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2649)
; FEATURE: misc feature
; NAME/KEY: (0)...(0)
; LOCATION: (0)...(0)
; OTHER INFORMATION: ORF 1255; Translation initiation factor
US-11-074-176-87

Query Match 3.5%; Score 35.4; DB 7; Length 2649;
Best Local Similarity 45.1%; Pred. No. 2.3;
Matches 132; Conservative 0; Mismatches 161; Indels 0; Gaps 0;
Qy 518 AAAAAAGACAAAAAGAAATCTGCACCTGGATGAATCATGAGATTCAAGAGGAAAGA 577
Db 431 AAGAAGGTTCAGCTTAATAGAGAGCTCAAAAAGCAAGAAAGAAATATCATGAGCATTTGA 490
Qy 578 AAAGAACTGCCGACACAGACTACTGGCTACAGCTGAAATTTATGTGAAAAATTATAACCA 637
Db 491 AGCATCTTAAAGAGAGCGTTCAGAAAAAGAAATATAGAAAAACAGTAAGAAAGTAATA 550
Qy 638 AGAACTCGGAGAGAAATATCATAGAAAAAGGCTATTGTTAAGGAAGTAATTGACAAAT 697
Db 551 AAAAAAGTAGAGCAACAGTGGAAAAAGAAAGTAATCGGACCAAAAAATCTTAAAAACCAT 610
Qy 698 ATACAGCTGTTGTAAGATGATTGATCTCGAGACAGCTGAAACTTGACCACTCAT 757
Db 611 CACCAGCCCGTTTAAAGAAAAATCAACCTGCAGATAATAAGGAAAAAGTTTACTACTCCTA 670
Qy 758 TAGACAGAGTAATTCACGACACACAGGAAAAAGAAATTCCTAGTTTAAATGGAGGC 810
Db 671 GAGTAACATATCTTGAAGACCAACCAAGGAGAAAAACGTGGTAAGGCTCGTGGC 723

RESULT 6

US-10-793-626-3726/c
; Sequence 3726, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:

; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P03480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3726
; LENGTH: 3043
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3726

Query Match 3.5%; Score 35; DB 1; Length 3043;
Best Local Similarity 43.7%; Pred. No. 3.2;
Matches 203; Conservative 0; Mismatches 260; Indels 2; Gaps 1;
Qy 517 AAAAAAGAGAAAAAGAAATCTGCACCTGGATGAATCATGAGATTGAGAGGAAAG 576
Db 1332 AATTAAGAGAAATTAGAGAAATGCTAATAAGATAAAATAATTTGTAGTAGAGAGTTGATAG 1273
Qy 577 AAAAGAACTGCCCGAACAGACTACTGGCTACAGCCTGAAATTTATGTGAAAAATTATAACC 636
Db 1272 ATATTATTAAGAACATGAATCTGAGATACAACTAGTTGAAGTTAATAATGAATTTGTTTT 1213
Qy 637 AAGAACTGGGAGAGAAATATATATAAGAAAAAGGCTATTGTTAAGGAAGTAATTGACAAA 696
Db 1212 GTTGATAGTATAGACAGTGTATCTTACGACAGAAAGGTTTCATAGAACTTTAAGAAATTC 1153
Qy 697 TATACAGCTGTTGTGAAGATGATTCTTGGAGACAAGCTGAAACTTTGACCACTCAT 756
Db 1152 ACAAGTTATTTTCTGATTTCACTAATTTATGTTTAAAGCGGATTTATCATCTTTGAA 1093
Qy 757 TTAGAGACAGTAATTCAGACACAGGAGAAAGAAATTTCTAGTTTAAATGGAGGCTACAGA 816
Db 1092 ATAGAAGATTTAGTTTTCAGAGCTGAAATTAAGCACTAGCTAAAGTATCGAAACCTGATATA 1033
Qy 817 GGAATGAAGGTACCCCTAGAAATCCATCAATGAGAAGACTTTTTTCAGCTACTATCGTCATT 876
Db 1032 TTAATAGATAATATAATAAAGTTTATTTAGTAAAAATGCTTATGCTATTATTAAGACAA 973
Qy 877 GAAACTGGCCCTTTAAAGGACGACAGATTGAAAGGAATTCAAATATGAAGACATTTCTAAA 936
Db 972 AGAATTGCTGGCTTAACAAATTAAGAAATGATTCAAGATGGCAAGAAAGAACTTTCAAT 913
Qy 937 CTGCTGAGTTTGAAG--ATTGTTTAAACAATACCTTTAAATCT 979
Db 912 TTTAATAATGTTATTAATCAAGAGATTACAAGACCATTTAAACCT 868

RESULT 7

US-10-793-626-4229
; Sequence 4229, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P03480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4229
; LENGTH: 3218
; TYPE: DNA
; ORGANISM: Artificial Sequence

; FEATURE: ; OTHER INFORMATION: Description of Artificial Sequence: synthetic ; OTHER INFORMATION: nucleic acid sequence US-10-793-626-4229									
Query Match 3.4%; Score 34; DB 1; Length 3218; Best Local Similarity 48.5%; Pred. No. 6.3; Mismatches 0; Indels 0; Gaps 0; Matches 94; Conservative 0									
QY	612	TGAATATTATGTGAAATATATACCAAGAACTGGGAGAGAAATATCATAGAAAAAGGC	671						
Db	3016	TGGAATAAGTCGATGTTTCATAGCATCGTCCAGTAGCAAGATCCACATGAATATCA	3075						
QY	672	TATTTGTAAGGAAGTAATTGACAAATATACAGCTGTTGTGAAGATGATTTCTGGAGA	731						
Db	3076	GGTTAAACCTAAAGATATTAAAGCATTAACAGATGCTGAGTTGTATTATTAATGTTTT	3135						
QY	732	CAAGCTGAACTTGACCACTCATTTAGACAGTAATCCAGCACAGGAAAAAGAAAT	791						
Db	3136	AAACCTAGAACTGGAATGGTTGGTTTGA AAAAGCACTTGACCAAGCAGGAAAAATCAAC	3195						
QY	792	TCTAGTTTTAAATG	805						
Db	3196	AAAAGATAAAATG	3209						
RESULT 8 US-10-793-626-3103 ; Sequence 3103, Application US/10793626 ; Publication No. US20050255478A1 ; GENERAL INFORMATION: ; APPLICANT: KIMMERLY, WILLIAM JOHN ; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS ; FILE REFERENCE: PU3480US ; CURRENT APPLICATION NUMBER: US/10/793,626 ; CURRENT FILING DATE: 2004-03-04 ; PRIOR APPLICATION NUMBER: 60/164,258 ; PRIOR FILING DATE: 1999-11-09 ; NUMBER OF SEQ ID NOS: 4472 ; SOFTWARE: PatentIn Ver. 2.1 ; SEQ ID NO 3103 ; LENGTH: 939 ; TYPE: DNA ; ORGANISM: Artificial Sequence ; FEATURE: ; OTHER INFORMATION: Description of Artificial Sequence: synthetic ; OTHER INFORMATION: nucleic acid sequence US-10-793-626-3103									
Query Match 3.3%; Score 33.4; DB 1; Length 939; Best Local Similarity 54.5%; Pred. No. 4.6; Matches 67; Conservative 0; Mismatches 56; Indels 0; Gaps 0;									
QY	743	TTGACCACTCATTTAGACAGTAATCCAGCACAGGAAAAAGAAATCTAGTTTTAA	802						
Db	71	TTGAGCATCAACATCCGGATACATTACTTCAAGAAGCTGAAATAAAAAATTCGTGCTGG	130						
QY	803	ATGGAGGCTACAGAGAAATGAGGTACCTAGAAATCCATCAATCCATGAGAAGACTTTTTCAG	862						
Db	131	AATCAATTTAGTGTCTAATGATCCGAATCAGATTTGATGATGTAATCAGCATGCAG	190						
QY	863	CTA 865							
Db	191	GTA 193							
RESULT 9 US-10-793-626-3369 ; Sequence 3369, Application US/10793626 ; Publication No. US20050255478A1 ; GENERAL INFORMATION: ; APPLICANT: KIMMERLY, WILLIAM JOHN ; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS									
Query Match 3.2%; Score 32.4; DB 7; Length 1167; Best Local Similarity 49.4%; Pred. No. 9.9; Matches 84; Conservative 0; Mismatches 86; Indels 0; Gaps 0;									
QY	513	GTCTAAAGAAAGAGAAAAAGAAATCTGCACCTGGATGAAATCATGAGATTTGAAGAGA	572						
Db	3	GTCTAAAGAAATGCTAGAGGCTTGCCTACATTTGAAAAAGAACAGGCGATTTGAACAGA	62						
; FILE REFERENCE: PU3480US ; CURRENT APPLICATION NUMBER: US/10/793,626 ; CURRENT FILING DATE: 2004-03-04 ; PRIOR APPLICATION NUMBER: 60/164,258 ; PRIOR FILING DATE: 1999-11-09 ; NUMBER OF SEQ ID NOS: 4472 ; SOFTWARE: PatentIn Ver. 2.1 ; SEQ ID NO 3369 ; LENGTH: 3148 ; TYPE: DNA ; ORGANISM: Artificial Sequence ; FEATURE: ; OTHER INFORMATION: Description of Artificial Sequence: synthetic ; OTHER INFORMATION: nucleic acid sequence US-10-793-626-3369									
Query Match 3.3%; Score 33.4; DB 1; Length 3148; Best Local Similarity 54.5%; Pred. No. 9.2; Matches 67; Conservative 0; Mismatches 56; Indels 0; Gaps 0;									
QY	743	TTGACCACTCATTTAGACAGTAATCCAGCACAGGAAAAAGAAATCTAGTTTTAA	802						
Db	705	TTGAGCATCAACATCCGGATACATTACTTCAAGAAGCTGAAATAAAAAATTCGTGCTGG	764						
QY	803	ATGGAGGCTACAGAGAAATGAGGTACCTAGAAATCCATCAATCCATGAGAAGACTTTTTCAG	862						
Db	765	AATCAATTTAGTGTCTAATGATCCGAATCAGATTTGATGATGTAATCAGCATGCAG	824						
QY	863	CTA 865							
Db	825	GTA 827							
RESULT 10 US-11-074-176-323 ; Sequence 323, Application US/11074176 ; Publication No. US20050250135A1 ; GENERAL INFORMATION: ; APPLICANT: Klaenhammer, Todd R. ; APPLICANT: Russell, William M. ; APPLICANT: Altermann, Eric ; APPLICANT: McAuliffe, Olivia ; APPLICANT: Perill, Andrea Azcarate ; TITLE OF INVENTION: Nucleic Acid Sequences Encoding ; TITLE OF INVENTION: Stress-Related Proteins and Uses Therefore ; FILE REFERENCE: 5051-694 ; CURRENT APPLICATION NUMBER: US/11/074,176 ; CURRENT FILING DATE: 2005-03-07 ; PRIOR APPLICATION NUMBER: 60/551,161 ; PRIOR FILING DATE: 2004-03-08 ; NUMBER OF SEQ ID NOS: 381 ; SOFTWARE: FastSeq for Windows Version 4.0 ; SEQ ID NO 323 ; LENGTH: 1167 ; TYPE: DNA ; ORGANISM: Lactobacillus acidophilus ; FEATURE: ; NAME/KEY: CDS ; LOCATION: (1)...(1167) ; FEATURE: ; NAME/KEY: misc_feature ; LOCATION: (0)...(0) ; OTHER INFORMATION: ORF 1259; N-utilization substance protein A US-11-074-176-323									
Query Match 3.2%; Score 32.4; DB 7; Length 1167; Best Local Similarity 49.4%; Pred. No. 9.9; Matches 84; Conservative 0; Mismatches 86; Indels 0; Gaps 0;									
QY	513	GTCTAAAGAAAGAGAAAAAGAAATCTGCACCTGGATGAAATCATGAGATTTGAAGAGA	572						
Db	3	GTCTAAAGAAATGCTAGAGGCTTGCCTACATTTGAAAAAGAACAGGCGATTTGAACAGA	62						

Qy 573 AAAGAAAAGAACTGCCGGAACAGACTACTGGCTACAGCCCTGAAATATTATTGTGAAAATTAT 632
Db 63 TGTATTGTTGATGCAATTAAAGCCGCTTTGGTAGCTGCATACAGAAAATTATATCA 122
Qy 633 AACCAAGAACTGGGAGAGAAATATCATAGAAAAGGCTATTGTTAAGG 682
Db 123 AGCAAAACGTTGAACTGATTTTGTATGAGAGAAAAGGCAATTTTAAAG 172

RESULT 11

US-11-074-176-91
; Sequence 91, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
; APPLICANT: Kleenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Peril, Andrea Ascarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; TITLE OF INVENTION: Stress-Related Proteins and Uses Therefore
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074,176
; PRIOR FILING DATE: 2005-03-07
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 91
; LENGTH: 1218
; TYPE: DNA
; ORGANISM: Lactobacillus acidophilus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1218)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: ORF 1259; N-utilization substance protein A
US-11-074-176-91

Query Match 3.2%; Score 32.4; DB 7; Length 1218;
Best Local Similarity 49.4%; Pred. No. 10;
Matches 84; Conservative 0; Mismatches 86; Indels 0; Gaps 0;
Qy 513 GTCTAAAGAAAGAGAAAGAAATCTGCACTGGATGAAATCATGAGATTGAGAGGA 572
Db 54 GTCTAAAGAAATGCTAGAGCGTTTCGCTACATTGGAAGAAAGAAAGGCGCATTTGAACAAGA 113
Qy 573 AAAGAAAAGAACTGCCGGAACAGACTACTGGCTACAGCCCTGAAATATTATTGTGAAAATTAT 632
Db 114 TGTATTGTTGATGCAATTAAAGCCGCTTTGGTAGCTGCATACAGAAAATTATATCA 173
Qy 633 AACCAAGAACTGGGAGAGAAATATCATAGAAAAGGCTATTGTTAAGG 682
Db 174 AGCAAAACGTTGAACTGATTTTGTATGAGAGAAAAGGCAATTTTAAAG 223

RESULT 12

US-10-793-626-4176/c
; Sequence 4176, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4176

; LENGTH: 3043
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4176

Query Match 3.2%; Score 32.4; DB 1; Length 3043;
Best Local Similarity 51.4%; Pred. No. 17;
Matches 75; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
Qy 836 AATCATCAATGAGAAGACTTTTTCAGCTACTATCTGTCATTGAACTGGCCCTTTAAAG 895
Db 2865 AATCTTGTAAATGTAATGAATCTATATCTACATTCATTTTGTATATGCTTAGAACGTG 2806
Qy 896 GACGAGAGTTGAAGGAATTCATATGAAGACATTTTCTAACTTGCTGAGTTGAAAT 955
Db 2805 TTGAAATAAATTCGTCAGTTCCATCTTCAGTCTATTTTAAATTTGTTAAATTCATCA 2746
Qy 956 TTGTTAAACAATACCTTTAAATCTTA 981
Db 2745 TTGTTAAATGTTCAAGAAAATGTTA 2720

RESULT 13

US-10-793-626-4417
; Sequence 4417, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4417
; LENGTH: 3078
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4417

Query Match 3.2%; Score 31.8; DB 1; Length 3078;
Best Local Similarity 44.1%; Pred. No. 25;
Matches 132; Conservative 0; Mismatches 167; Indels 0; Gaps 0;
Qy 596 ACTACTGGCTACAGCCCTGAAATATTGTTGAAAATTATTAACCAAGAACTGGAGAGAAAT 655
Db 660 ACTACTTGTATTAGATTCTTATTGTTAAATTTAACTATAAAACACAGGCTGCTGAAG 719
Qy 656 ATCATAGAAAAGGCTATTGTTAAGGAAGTAATTCAGAAATATACAGCTGTTGCTGAAGA 715
Db 720 ATAAAAAATCATCTGTAACAACTACATCGGCTAGTCAATCGCCATTCGATGTTTGAAG 779
Qy 716 TGATTGATCTGGAGACAAGCTGAAACTTCACAGACTCATTTTAGAGACAGTAATTCAG 775
Db 780 CAATGGGTGTAAGAAAATATTAAACATCTAGATGCTTGTATTACTAGACTAGAGTAG 839
Qy 776 CACCAGAAAAGAAATTCATGTTTAAATCGAGGCTACAGAGGAATGAAGGTACCCCTAG 835
Db 840 AAGTAATGAAAATCAAAAGTTGATGTTGCGGCCCTTAAATCATTAGGTGCTTCAGGAG 899
Qy 836 AATCCATCAATGAGAGACTTTTTCAGCTACTACTCGTCATTGAACTGGCCCTTTAAA 894
Db 900 TACTTGAAGTCGGTAACAATATGCAAGCTATTTTCGGTCTCTAAATCTGATCAATCA 958

RESULT 14
US-10-658-986-5/c
; Sequence 5, Application US/10658986
; Publication No. US20050255476A1
; GENERAL INFORMATION:
; APPLICANT: Clark, Abbot F.
; TITLE OF INVENTION: Methods for Diagnosing Glaucoma and Discovering Anti-Glaucoma Drugs
; FILE REFERENCE: 1581 US F
; CURRENT APPLICATION NUMBER: US/10/658,986
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: PCT/US97/21054
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: USSN 60/033,227
; PRIOR FILING DATE: 1996-12-05
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 125594
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-658-986-5

Query Match 3.2%; Score 31.8; DB 1; Length 125594;
Best Local Similarity 46.9%; Pred. No. 1.9e+02;
Matches 99; Conservative 0; Mismatches 112; Indels 0; Gaps 0;
QY 497 CCAGAGCTCAACTCAGTCTAAAGAAAAGAAAGAAAGAAATCTGCACCTGGATGAAATCA 556
DB 38437 CACTGTGCCACCTAGGCTGTAAATGACATATAAATCTCGTGGCAACCACTAAGAA 38378
QY 557 TGGAGATTGAAGAGGAAAAGAAAAGAACTGCCGAAACAGACTACTGGCTACAGCCCTGAAA 616
DB 38377 AATACTAAAACATACAAAGAAAGAAATGACAAAGAAATTAATAATCGTTACACTAGAAA 38318
QY 617 TTATTGTGAATTTAATCAACAGAACTGGGAGAGAAATATCATAGAAAAGGCTATTG 676
DB 38317 ATATGTPAATACAAAGAAAGAAAGGTAGTAAATGGAAGATTTTAAGAAACAAAGAGACACAAA 38258
QY 677 TTAAGGAAGTAATTGACAAATATACAGCTGT 707
DB 38257 TAAAGGAACAAATGCGCAAAATATCAGAAGT 38227

RESULT 15
US-10-467-962B-102
; Sequence 102, Application US/10467962B
; Publication No. US20050246784A1
; GENERAL INFORMATION:
; APPLICANT: Plesch, Gunmar
; APPLICANT: Blau, Astrid
; APPLICANT: Daeschner, Klaus
; APPLICANT: Klein, Mathieu
; TITLE OF INVENTION: Identification of Herbicidally Active Substances
; FILE REFERENCE: 2000 857
; CURRENT APPLICATION NUMBER: US/10/467,962B
; CURRENT FILING DATE: 2003-08-14
; PRIOR APPLICATION NUMBER: PCT/EP02/01466
; PRIOR FILING DATE: 2002-02-13
; NUMBER OF SEQ ID NOS: 109
; SOFTWARE: PatentIn Vers. 2.0
; SEQ ID NO 102
; LENGTH: 2361
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2361)
; OTHER INFORMATION: ORF F4F7.26, Linie 246473
US-10-467-962B-102

Query Match 3.1%; Score 31.4; DB 1; Length 2361;
Best Local Similarity 51.0%; Pred. No. 28;

Matches 74; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
QY 601 TGGCTACAGCCCTGAAATTAATTGTGAAAAATTATAACCAAGAAACTGGGAGAGAAAAATATCAT 660
DB 490 TGGGCCGATCCTAAGATTGTTAAAGAACTCATAGATGAGAAAAATGATGAGCTGCTTGGT 549
QY 661 AAGAAAAAGGCTATTGTTAAAGAAAGTAATTGACAAATATACAGCTGTTGTGAAGATGATT 720
DB 550 GAGAAAAACCGCTGCTGATAATGAAAAACCTACAAAAAGAGGAGAGAAAGGAGAAAGCCT 609
QY 721 GATTCTGGAGACAAGCTGAAACTTG 745
DB 610 GCCAAAGTCGAGGAAAAGAAAGCTG 634

Search completed: November 27, 2005, 02:44:30
Job time : 354.647 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 12:53:47 ; Search time 116,932 Seconds
(without alignments)
10694.773 Million cell updates/sec

Title: US-09-555-529-18
Perfect score: 22
Sequence: 1 agaaagtgcgtgcgtggt 22

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.1*

- 1: gb_ba.*
- 2: gb_in.*
- 3: gb_env.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pi.*
- 9: gb_ro.*
- 10: gb_sts.*
- 11: gb_sy.*
- 12: gb_un.*
- 13: gb_vi.*
- 14: gb_htg.*
- 15: gb_pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	100.0	22	6	AX003325 Sequence
2	22	100.0	1002	6	AX003310 Sequence
3	22	100.0	1117	6	CQ722054 Sequence
4	22	100.0	1528	6	AX003308 Sequence
5	22	100.0	1528	8	HSBJ273
6	22	100.0	23065	8	HUMATP5G
7	22	100.0	174361	8	AL158044
8	22	100.0	194563	14	AL391687
9	18.4	83.6	7111	14	AC017711
10	18.4	83.6	166348	14	CR847851
11	18.4	83.6	168906	5	AC146480
12	18.4	83.6	170597	5	AL929558
13	18.4	83.6	177953	2	AC018491
14	18.4	83.6	297848	2	AE002611
15	17.8	80.9	460	1	AY525807
16	17.8	80.9	889	5	EX931016
17	17.8	80.9	2169	5	CR387275
18	17.8	80.9	6584	3	AY236226

C	19	17.8	80.9	10570	1	AE005005
C	20	17.8	80.9	101798	14	AC150077
C	21	17.8	80.9	110000	1	CP000075_08
C	22	17.4	79.1	746	1	ECOAMS
C	23	17.4	79.1	1613	1	ECU49647
C	24	17.4	79.1	3661	6	BD247478
C	25	17.4	79.1	3661	6	AR408798
C	26	17.4	79.1	4000	1	ECGAMS
C	27	17.4	79.1	4463	5	AF061799
C	28	17.4	79.1	14663	1	D90744
C	29	17.4	79.1	103920	14	AC015817_3
C	30	17.4	79.1	110000	1	U00096_11
C	31	17.4	79.1	110000	15	AP008216_222
C	32	17.4	79.1	172427	15	AC060755
C	33	17.4	79.1	175570	9	AL844893
C	34	17.4	79.1	222949	14	AC163667
C	35	17.4	79.1	300029	15	AE017121
C	36	17.2	78.2	412	6	CQ059049
C	37	17.2	78.2	412	6	CQ078469
C	38	17.2	78.2	412	6	CQ109433
C	39	17.2	78.2	412	6	CQ148085
C	40	17.2	78.2	412	6	CQ183469
C	41	17.2	78.2	412	6	CQ207864
C	42	17.2	78.2	412	6	CQ231346
C	43	17.2	78.2	412	6	CQ269518
C	44	17.2	78.2	412	6	CQ306575
C	45	17.2	78.2	412	6	CQ343626

ALIGNMENTS

RESULT 1

AX003325

LOCUS

DEFINITION

AX003325

ACCESSION

AX003325.1

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

Location/Qualifiers

1. .22

/organism="Homo sapiens"

/mol_type="unassigned DNA"

/db_xref="taxon:9606"

ORIGIN

Query Match

Best Local Similarity

Matches

Qy

Db

RESULT 2

AX003310

LOCUS

DEFINITION

AX003310

ACCESSION

VERSION

KEYWORDS

AX003310

Sequence 3 from Patent WO9929845.

AX003310.1

GI:9927127

linear

DNA

1002 bp

Sequence 3 from Patent WO9929845.

AX003310

AX003310.1

GI:9927127

linear

DNA

1002 bp

Sequence 3 from Patent WO9929845.

AX003310

AX003310.1

GI:9927127

linear

DNA

1002 bp

Sequence 3 from Patent WO9929845.

AX003310

AX003310.1

GI:9927127

linear

DNA

1002 bp

Sequence 3 from Patent WO9929845.

AX003310

AX003310.1

GI:9927127

linear

DNA

1002 bp

Sequence 3 from Patent WO9929845.

AX003310

AX003310.1

GI:9927127

linear

DNA

1002 bp

Sequence 3 from Patent WO9929845.

AX003310

AX003310.1

GI:9927127

linear

DNA

1002 bp

Sequence 3 from Patent WO9929845.

AX003310

AX003310.1

GI:9927127

linear

DNA

1002 bp

Sequence 3 from Patent WO9929845.

AX003310

AX003310.1

GI:9927127

linear

DNA

1002 bp

Sequence 3 from Patent WO9929845.

AX003310

AX003310.1

GI:9927127

linear

DNA

1002 bp

Sequence 3 from Patent WO9929845.

AX003310

AX003310.1

GI:9927127

linear

DNA

1002 bp

Sequence 3 from Patent WO9929845.

AX003310

AX003310.1

GI:9927127

linear

DNA

1002 bp

Sequence 3 from Patent WO9929845.

AX003310

AX003310.1

GI:9927127

linear

DNA

1002 bp

Sequence 3 from Patent WO9929845.

AX003310

AX003310.1

GI:9927127

linear

DNA

1002 bp

Sequence 3 from Patent WO9929845.

AX003310

AX003310.1

GI:9927127

linear

DNA

1002 bp

Sequence 3 from Patent WO9929845.

AX003310

AX003310.1

GI:9927127

linear

DNA

1002 bp

Sequence 3 from Patent WO9929845.

AX003310

AX003310.1

GI:9927127

linear

DNA

1002 bp

Sequence 3 from Patent WO9929845.

AX003310

AX003310.1

GI:9927127

linear

DNA

1002 bp

Sequence 3 from Patent WO9929845.

AX003310

AX003310.1

GI:9927127

linear

DNA

1002 bp

Sequence 3 from Patent WO9929845.

AX003310

AX003310.1

GI:9927127

linear

DNA

1002 bp

Sequence 3 from Patent WO9929845.

AX003310

AX003310.1

GI:9927127

linear

DNA

1002 bp

Sequence 3 from Patent WO9929845.

AX003310

AX003310.1

GI:9927127

linear

DNA

1002 bp

Sequence 3 from Patent WO9929845.

AX003310

AX003310.1

GI:9927127

linear

DNA

1002 bp

Sequence 3 from Patent WO9929845.

AX003310

AX003310.1

GI:9927127

linear

DNA

1002 bp

Sequence 3 from Patent WO9929845.

AX003310

AX003310.1

GI:9927127

linear

DNA

1002 bp

Sequence 3 from Patent WO9929845.

AX003310

AX003310.1

GI:9927127

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1
AUTHORS Angulo-Mora, J.F. and Mauffrey, P.
TITLE Sequences coding for kin17 protein and their applications
JOURNAL Patent: WO 9929845-A 3 17-JUN-1999;
ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE (FR)
FEATURES source Location/Qualifiers
1..1002
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 22; DB 6; Length 1002;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGAAAGTGATCGTGCCTGGT 22
|||||
Db 32 AGAAAGTGATCGTGCCTGGT 53
|||||
RESULT 3
LOCUS CQ722054 1117 bp DNA linear PAT 03-FEB-2004
DEFINITION Sequence 7988 from Patent WO02068579.
ACCESSION CQ722054
VERSION CQ722054.1 GI:42282911
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1
AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
TITLE Kits, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL Patent: WO 02068579-A 7988 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES source Location/Qualifiers
1..1117
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 22; DB 6; Length 1117;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGAAAGTGATCGTGCCTGGT 22
|||||
Db 67 AGAAAGTGATCGTGCCTGGT 88
|||||
RESULT 4
LOCUS AX003308 1528 bp DNA linear PAT 24-AUG-2000
DEFINITION Sequence 1 from Patent WO9929845.
ACCESSION AX003308
VERSION AX003308.1 GI:9927125
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.
REFERENCE 1
AUTHORS Angulo-Mora, J.F. and Mauffrey, P.
TITLE Sequences coding for kin17 protein and their applications
JOURNAL Patent: WO 9929845-A 1 17-JUN-1999;
ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE (FR)
FEATURES source Location/Qualifiers
1..1528
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 22; DB 6; Length 1528;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGAAAGTGATCGTGCCTGGT 22
|||||
Db 40 AGAAAGTGATCGTGCCTGGT 61
|||||
RESULT 5
LOCUS HSAJ5273 1528 bp mRNA linear PRI 13-DEC-2000
DEFINITION Homo sapiens mRNA for Kin17 protein.
ACCESSION AJ005273
VERSION AJ005273.1 GI:3850703
KEYWORDS KIN17 gene.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1
AUTHORS Kannouche, P., Mauffrey, P., Pinon-Lataillade, G., Mattei, M.G.,
Sarasin, A., Daya-Grosjean, L. and Angulo, J.F.
TITLE Molecular cloning and characterization of the human KIN17 cDNA
encoding a component of the UVC response that is conserved among
metazoans
JOURNAL Carcinogenesis 21 (9), 1701-1710 (2000)
PUBMED 10984102
REFERENCE 2
AUTHORS Mauffrey, P.
TITLE Direct Submission
JOURNAL Submitted (08-APR-1998) Mauffrey P., LGR, CEA-DSV-DRR, BP 6 -
Fontenay aux Roses Cedex, 92265, FRANCE
REMARK Revised by [3]
REFERENCE 3 (bases 1 to 1528)
AUTHORS Mauffrey, P.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1998) Mauffrey P., LGR, CEA-DSV-DRR, BP 6 -
Fontenay aux Roses Cedex, 92265, FRANCE
COMMENT On Nov 7, 1998 this sequence version replaced gi:3046739.
Kannouche P., Mauffrey P., Pinon-Lataillade G., Biard D.S.F.,
Angulo J.F.
Sequences codant pour la Proteine kin17 et leurs Applications -
French Patent Nr 97 15536 1:1-50(1997).
FEATURES source Location/Qualifiers
1..1528
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/sex="male"
/tissue type="testis"
1..1528
/gene="Kin17"
66..1247
/codon_start=1
/protein_id="CAA06462.1"
/db_xref="GI:3850704"

```

/db|xref="GOA:O60870"
/db|xref="UniProt/T/EMBL:O60870"
/translation="MGKSDFTTPRAIANRIKSGLOKLRWYQCMQCKQCRDENGFKCH
CMSSHQRLLASSENQQPMDFSEFRNDFLELLRRPGTKRVHNNIVNYEYISHR
OHLMNATQWETLTDFTKWLGREGCKVDPTFGWYIQYIDRDPETIRQLELEKKKK
QDLDEBKAKFIEQVRRLGEGKEVPTFTLSRENDEKVTNLSKGCSSSGAT
SSKSTLPGSALKITIGSSASVKKRQSSSTQSKKKKKSALDEIMEIEEKRTAR
TDYWLQPEIIVKIIITKLGKRYHKKKAIKVEVIDKYTAVMKIDSGDKLKDQTHLET
VIPAPGRKILVLNGYRGNGECTLESINEKFTSATIVTETGPKRRRVEGIGQYEDISK
A"
ORIGIN
Query Match 100.0%; Score 22; DB 8; Length 1528;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AGAAAGTCATCGTCGCGTGGT 22
Db 40 AGAAAGTCATCGTCGCGTGGT 61
RESULT 6
HUMATPSG/c
LOCUS HUMATPSG 23065 bp DNA linear PRI 13-FEB-2003
DEFINITION Homo sapiens gene for ATP synthase gamma-subunit, complete cds.
ACCESSION D16561
VERSION D16561.1 GI:468446
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE
1 (bases 1 to 23065)
Matsuda,C., Endo,H., Ohta,S. and Kagawa,Y.
Gene structure of human mitochondrial ATP synthase gamma-subunit.
Tissue specificity produced by alternative RNA splicing
J. Biol. Chem. 268 (33), 24950-24958 (1993)
PUBMED 8227057
REFERENCE
2 (bases 1 to 23065)
Kagawa,Y.
Direct Submission
Submitted (23-JUN-1993) Yasuo Kagawa, Jichi Medical School,
Department of Biochemistry, 311-1 Yakushiji, Minamikawachi-machi,
Tochigi 329-0498, Japan (E-mail:ykagawa@ddj.nig.ac.jp,
Tel:81-285-44-2111(ex.3149), Fax:81-285-44-1827)
LOCATION/Qualifiers
1. |23065
|organism="Homo sapiens"
|mol_type="genomic DNA"
|db|xref="taxon:9606"
|tissue_type="blood"
|clone_lib="lambda dash II"
|note="clones hATPG[2, 5, 21 and 23]"
4188..4274
/number=1
Join(4219..4274,11169..11203,12114..12245,14057..14261,
14839..14982,15094..15158,17329..17484,17817..17913,
22713..22716)
|note="H(heart)-type ATP synthase gamma-subunit"
/codon_start=1
/product="ATP synthase gamma-subunit"
/protein_id="BAA03994.1"
/db|xref="GI:665584"
/translation="MFSRAGVAGLSAWTLQPOWIOVRNMTLKDITRLKSIKNIOKI
TKSMKMAAKYARAEKLPARIYGLGSLALYERADIKGPEDKKHLLIGVSSDRGL
CGAIISSIAKQMSKSEVATLTAAQKVMVLGVGDKIRGILYRTHSDQFLVAFKVRKRP
PTFGDASVIALELLNSGYEFDGSIIFNKRFSVISYKTEKPIFSLNTVADSMSIY
DDIDADVLQNYQEVNLANIYYSLKSTTSQSRMTAMDNASKNASEMIDKLTFTN
RTRQAVITKELIEIISGAAL"
4188..4274
/number=1
Join(4219..4274,11169..11203,12114..12245,14057..14261,
14839..14982,15094..15158,17329..17484,17817..17913,
22028..22034)
sequence.
AL158044
ACCESSION AL158044
LOCUS AL158044/c
DEFINITION Human DNA sequence from clone RP11-264C14 on chromosome 10 Contains
the 5' end of the gene for a novel protein (MGC10848), the ITH2
gene for inter-alpha (globulin) inhibitor, H2 polypeptide (H2P),
the KIN gene for KIN, antigenic determinant of reca protein homolog
(mouse), the 3' end of the ATP5C1 gene for ATP synthase, H+
transporting, mitochondrial F1 complex, gamma polypeptide 1
(ATP5C, ATP5C1), a novel gene and two CpG islands, complete
sequence.
AL158044

```

```

VERSION      AL158044.17  GI:16214543
KEYWORDS     HTG; ATP5C; ATP5C1; ATP5C11; CpG island; H2P; H2IH2; KIN; MGC10848.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Euarchontoglires; Primates; Catarrhini,
Homnidae; Homo.
REFERENCE    1 (bases 1 to 174361)
AUTHORS      Lovell, J.
TITLE        Direct Submission
JOURNAL      Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
COMMENT      Clone requests: clonerequest@sanger.ac.uk
On Oct 17, 2001 this sequence version replaced gi:15131961.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
En: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
RP11-264C14 is from the library RPCT-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vegas@sanger.ac.uk
-----
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one subclone; and the assembly was confirmed by restriction digest,
except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source            1..174361
                     1.174361
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-264C14"
                     /clone_lib="RPCT-11.1"
     misc_feature      1
                     /notes="Clone left end: RP11-264C14"
     gene              join(complement(3206..3285),
                     complement(AL355374.14:26760..81619))
                     /gene="ITIHS"
     mRNA              /locus_tag="RP11-676E7.1-003"
                     join(complement(3206..3285),
                     complement(AL355374.14:81450..81619),
                     complement(AL355374.14:80319..80435),
                     complement(AL355374.14:50238..50406),
                     complement(AL355374.14:44092..44401),
                     complement(AL355374.14:40790..41349),
                     complement(AL355374.14:37479..37532),
                     complement(AL355374.14:34005..34121),
                     complement(AL355374.14:30367..30744),
                     complement(AL355374.14:23606..27721))
                     /gene="ITIHS"
     CDS               /locus_tag="RP11-676E7.1-003"
                     /product="inter-alpha (globulin) inhibitor H5"
                     /notes="match: ESTs: BE386984.1 BE387282.1 BG333950.1
                     BG476698.1 BQ225649.1
                     match: cDNAs: BC004282.1"
                     join(complement(3206..3215),
                     complement(AL355374.14:81450..81619),
                     complement(AL355374.14:81450..81619),
                     complement(AL355374.14:80319..80435)),
                     complement(AL355374.14:26760..81619))
complement(AL355374.14:50238..50406),
complement(AL355374.14:44092..44401),
complement(AL355374.14:40790..41349),
complement(AL355374.14:37479..37532),
complement(AL355374.14:34005..34121),
complement(AL355374.14:30367..30744),
complement(AL355374.14:23606..27721))
/gene="ITIHS"
/locus_tag="RP11-676E7.1-003"
/product="inter-alpha (globulin) inhibitor H5"
/notes="match: ESTs: AAL34750.1 AA367857.1 AA367892.1
A1089902.1 A1090076.1 A1339610.1 A1417871.1 A14079707.1
AL540791.1 AL543337.1 AL544674.1 AL544702.1 AL550125.1
AL550297.1 AL550915.1 AL554626.1 AL5548595.1 AL5569892.1
AL5570737.1 AL574092.1 AL574207.1 AL574544.1 AL575373.1
AL577265.1 AW473748.1 BE893152.1 BE896223.1 BE898209.1
BF677721.1 BF689672.1 BF690566.1 BF726069.1 BF998687.1
BG006306.1 BG006608.1 BG282300.1 BG282310.1 BG286634.1
BG422562.1 BG435601.1 BG478666.1 BG676410.1 BG701677.1
BG767953.1 B104702.1 B1553925.1 B1559648.1 BM016522.1
BM542130.1 BM542378.1 BM549311.1 BM557411.1 BM675641.1
BM675641.1
complement(25552..25715), complement(24379..24480),
complement(20853..21103),
complement(AL355374.14:81450..81619),
complement(AL355374.14:80319..80435),
complement(AL355374.14:50238..50406),
complement(AL355374.14:44092..44401),
complement(AL355374.14:40790..41349),
complement(AL355374.14:37479..37532),
complement(AL355374.14:34005..34121),
complement(AL355374.14:30367..30744),
complement(AL355374.14:23606..27721))
/gene="ITIHS"
/locus_tag="RP11-676E7.1-001"
join(complement(50428..50596), complement(39256..39300),
complement(25552..25715), complement(24379..24480),
complement(20853..21103),
complement(AL355374.14:81450..81619),
complement(AL355374.14:80319..80435),
complement(AL355374.14:50238..50406),
complement(AL355374.14:44092..44401),
complement(AL355374.14:40790..41349),
complement(AL355374.14:37479..37532),
complement(AL355374.14:34005..34121),
complement(AL355374.14:30367..30744),
complement(AL355374.14:23606..27721))
/gene="ITIHS"
/locus_tag="RP11-676E7.1-001"
/product="inter-alpha (globulin) inhibitor H5"
/notes="match: ESTs: AAL34750.1 AA367857.1 AA367892.1
A1089902.1 A1090076.1 A1339610.1 A1417871.1 A14079707.1
AL540791.1 AL543337.1 AL544674.1 AL544702.1 AL550125.1
AL550297.1 AL550915.1 AL554626.1 AL5548595.1 AL5569892.1
AL5570737.1 AL574092.1 AL574207.1 AL574544.1 AL575373.1
AL577265.1 AW473748.1 BE893152.1 BE896223.1 BE898209.1
BF677721.1 BF689672.1 BF690566.1 BF726069.1 BF998687.1
BG006306.1 BG006608.1 BG282300.1 BG282310.1 BG286634.1
BG422562.1 BG435601.1 BG478666.1 BG676410.1 BG701677.1
BG767953.1 B104702.1 B1553925.1 B1559648.1 BM016522.1
BM542130.1 BM542378.1 BM549311.1 BM557411.1 BM675641.1
BM675641.1
complement(50428..50596), complement(39256..39300),
complement(25552..25715), complement(24379..24480),
complement(20853..21103),
complement(AL355374.14:81450..81619),
complement(AL355374.14:80319..80435),
complement(AL355374.14:50238..50406),
complement(AL355374.14:44092..44401),
complement(AL355374.14:40790..41349),
complement(AL355374.14:37479..37532),
complement(AL355374.14:34005..34121),
complement(AL355374.14:30367..30744),
complement(AL355374.14:23606..27721))
/gene="ITIHS"
/locus_tag="RP11-676E7.1-001"
/product="inter-alpha (globulin) inhibitor H5"
/notes="match: ESTs: AAL34750.1 AA367857.1 AA367892.1
A1089902.1 A1090076.1 A1339610.1 A1417871.1 A14079707.1
AL540791.1 AL543337.1 AL544674.1 AL544702.1 AL550125.1
AL550297.1 AL550915.1 AL554626.1 AL5548595.1 AL5569892.1
AL5570737.1 AL574092.1 AL574207.1 AL574544.1 AL575373.1
AL577265.1 AW473748.1 BE893152.1 BE896223.1 BE898209.1
BF677721.1 BF689672.1 BF690566.1 BF726069.1 BF998687.1
BG006306.1 BG006608.1 BG282300.1 BG282310.1 BG286634.1
BG422562.1 BG435601.1 BG478666.1 BG676410.1 BG701677.1
BG767953.1 B104702.1 B1553925.1 B1559648.1 BM016522.1
BM542130.1 BM542378.1 BM549311.1 BM557411.1 BM675641.1
BM675641.1
complement(50428..50596), complement(39256..39300),
complement(25552..25715), complement(24379..24480),
complement(20853..21103),
complement(AL355374.14:81450..81619),
complement(AL355374.14:80319..80435),
complement(AL355374.14:50238..50406),
complement(AL355374.14:44092..44401),
complement(AL355374.14:40790..41349),
complement(AL355374.14:37479..37532),
complement(AL355374.14:34005..34121),
complement(AL355374.14:30367..30744),
complement(AL355374.14:23606..27721))
/gene="ITIHS"
/locus_tag="RP11-676E7.1-001"
/product="inter-alpha (globulin) inhibitor H5"
/notes="match: ESTs: AAL34750.1 AA367857.1 AA367892.1
A1089902.1 A1090076.1 A1339610.1 A1417871.1 A14079707.1
AL540791.1 AL543337.1 AL544674.1 AL544702.1 AL550125.1
AL550297.1 AL550915.1 AL554626.1 AL5548595.1 AL5569892.1
AL5570737.1 AL574092.1 AL574207.1 AL574544.1 AL575373.1
AL577265.1 AW473748.1 BE893152.1 BE896223.1 BE898209.1
BF677721.1 BF689672.1 BF690566.1 BF726069.1 BF998687.1
BG006306.1 BG006608.1 BG282300.1 BG282310.1 BG286634.1
BG422562.1 BG435601.1 BG478666.1 BG676410.1 BG701677.1
BG767953.1 B104702.1 B1553925.1 B1559648.1 BM016522.1
BM542130.1 BM542378.1 BM549311.1 BM557411.1 BM675641.1
BM675641.1
complement(50428..50596), complement(39256..39300),
complement(25552..25715), complement(24379..24480),
complement(20853..21103),
complement(AL355374.14:81450..81619),
complement(AL355374.14:80319..80435),
complement(AL355374.14:50238..50406),
complement(AL355374.14:44092..44401),
complement(AL355374.14:40790..41349),
complement(AL355374.14:37479..37532),
complement(AL355374.14:34005..34121),
complement(AL355374.14:30367..30744),
complement(AL355374.14:23606..27721))
/gene="ITIHS"
/locus_tag="RP11-676E7.1-001"
/product="inter-alpha (globulin) inhibitor H5"
/notes="match: ESTs: AAL34750.1 AA367857.1 AA367892.1
A1089902.1 A1090076.1 A1339610.1 A1417871.1 A14079707.1
AL540791.1 AL543337.1 AL544674.1 AL544702.1 AL550125.1
AL550297.1 AL550915.1 AL554626.1 AL5548595.1 AL5569892.1
AL5570737.1 AL574092.1 AL574207.1 AL574544.1 AL575373.1
AL577265.1 AW473748.1 BE893152.1 BE896223.1 BE898209.1
BF677721.1 BF689672.1 BF690566.1 BF726069.1 BF998687.1
BG006306.1 BG006608.1 BG282300.1 BG282310.1 BG286634.1
BG422562.1 BG435601.1 BG478666.1 BG676410.1 BG701677.1
BG767953.1 B104702.1 B1553925.1 B1559648.1 BM016522.1
BM542130.1 BM542378.1 BM549311.1 BM557411.1 BM675641.1
BM675641.1
complement(50428..50596), complement(39256..39300),
complement(25552..25715), complement(24379..24480),
complement(20853..21103),
complement(AL355374.14:81450..81619),
complement(AL355374.14:80319..80435),
complement(AL355374.14:50238..50406),
complement(AL355374.14:44092..44401),
complement(AL355374.14:40790..41349),
complement(AL355374.14:37479..37532),
complement(AL355374.14:34005..34121),
complement(AL355374.14:30367..30744),
complement(AL355374.14:23606..27721))
/gene="ITIHS"
/locus_tag="RP11-676E7.1-001"
/product="inter-alpha (globulin) inhibitor H5"
/notes="match: ESTs: AAL34750.1 AA367857.1 AA367892.1
A1089902.1 A1090076.1 A1339610.1 A1417871.1 A14079707.1
AL540791.1 AL543337.1 AL544674.1 AL544702.1 AL550125.1
AL550297.1 AL550915.1 AL554626.1 AL5548595.1 AL5569892.1
AL5570737.1 AL574092.1 AL574207.1 AL574544.1 AL575373.1
AL577265.1 AW473748.1 BE893152.1 BE896223.1 BE898209.1
BF677721.1 BF689672.1 BF690566.1 BF726069.1 BF998687.1
BG006306.1 BG006608.1 BG282300.1 BG282310.1 BG286634.1
BG422562.1 BG435601.1 BG478666.1 BG676410.1 BG701677.1
BG767953.1 B104702.1 B1553925.1 B1559648.1 BM016522.1
BM542130.1 BM542378.1 BM549311.1 BM557411.1 BM675641.1
BM675641.1
complement(50428..50596), complement(39256..39300),
complement(25552..25715), complement(24379..24480),
complement(20853..21103),
complement(AL355374.14:81450..81619),
complement(AL355374.14:80319..80435),
complement(AL355374.14:50238..50406),
complement(AL355374.14:44092..44401),
complement(AL355374.14:40790..41349),
complement(AL355374.14:37479..37532),
complement(AL355374.14:34005..34121),
complement(AL355374.14:30367..30744),
complement(AL355374.14:23606..27721))
/gene="ITIHS"
/locus_tag="RP11-676E7.1-001"
/product="inter-alpha (globulin) inhibitor H5"
/notes="match: ESTs: AAL34750.1 AA367857.1 AA367892.1
A1089902.1 A1090076.1 A1339610.1 A1417871.1 A14079707.1
AL540791.1 AL543337.1 AL544674.1 AL544702.1 AL550125.1
AL550297.1 AL550915.1 AL554626.1 AL5548595.1 AL5569892.1
AL5570737.1 AL574092.1 AL574207.1 AL574544.1 AL575373.1
AL577265.1 AW473748.1 BE893152.1 BE896223.1 BE898209.1
BF677721.1 BF689672.1 BF690566.1 BF726069.1 BF998687.1
BG006306.1 BG006608.1 BG282300.1 BG282310.1 BG286634.1
BG422562.1 BG435601.1 BG478666.1 BG676410.1 BG701677.1
BG767953.1 B104702.1 B1553925.1 B1559648.1 BM016522.1
BM542130.1 BM542378.1 BM549311.1 BM557411.1 BM675641.1
BM675641.1
complement(50428..50596), complement(39256..39300),
complement(25552..25715), complement(24379..24480),
complement(20853..21103),
complement(AL355374.14:81450..81619),
complement(AL355374.14:80319..80435),
complement(AL355374.14:50238..50406),
complement(AL355374.14:44092..44401),
complement(AL355374.14:40790..41349),
complement(AL355374.14:37479..37532),
complement(AL355374.14:34005..34121),
complement(AL355374.14:30367..30744),
complement(AL355374.14:23606..27721))
/gene="ITIHS"
/locus_tag="RP11-676E7.1-001"
/product="inter-alpha (globulin) inhibitor H5"
/notes="match: ESTs: AAL34750.1 AA367857.1 AA367892.1
A1089902.1 A1090076.1 A1339610.1 A1417871.1 A14079707.1
AL540791.1 AL543337.1 AL544674.1 AL544702.1 AL550125.1
AL550297.1 AL550915.1 AL554626.1 AL5548595.1 AL5569892.1
AL5570737.1 AL574092.1 AL574207.1 AL574544.1 AL575373.1
AL577265.1 AW473748.1 BE893152.1 BE896223.1 BE898209.1
BF677721.1 BF689672.1 BF690566.1 BF726069.1 BF998687.1
BG006306.1 BG006608.1 BG282300.1 BG282310.1 BG286634.1
BG422562.1 BG435601.1 BG478666.1 BG676410.1 BG701677.1
BG767953.1 B104702.1 B1553925.1 B1559648.1 BM016522.1
BM542130.1 BM542378.1 BM549311.1 BM557411.1 BM675641.1
BM675641.1
complement(50428..50596), complement(39256..39300),
complement(25552..25715), complement(24379..24480),
complement(20853..21103),
complement(AL355374.14:81450..81619),
complement(AL355374.14:80319..80435),
complement(AL355374.14:50238..50406),
complement(AL355374.14:44092..44401),
complement(AL355374.14:40790..41349),
complement(AL355374.14:37479..37532),
complement(AL355374.14:34005..34121),
complement(AL355374.14:30367..30744),
complement(AL355374.14:23606..27721))
/gene="ITIHS"
/locus_tag="RP11-676E7.1-001"
/product="inter-alpha (globulin) inhibitor H5"
/notes="match: ESTs: AAL34750.1 AA367857.1 AA367892.1
A1089902.1 A1090076.1 A1339610.1 A1417871.1 A14079707.1
AL540791.1 AL543337.1 AL544674.1 AL544702.1 AL550125.1
AL550297.1 AL550915.1 AL554626.1 AL5548595.1 AL5569892.1
AL5570737.1 AL574092.1 AL574207.1 AL574544.1 AL575373.1
AL577265.1 AW473748.1 BE893152.1 BE896223.1 BE898209.1
BF677721.1 BF689672.1 BF690566.1 BF726069.1 BF998687.1
BG006306.1 BG006608.1 BG282300.1 BG282310.1 BG286634.1
BG422562.1 BG435601.1 BG478666.1 BG676410.1 BG701677.1
BG767953.1 B104702.1 B1553925.1 B1559648.1 BM016522.1
BM542130.1 BM542378.1 BM549311.1 BM557411.1 BM675641.1
BM675641.1
complement(50428..50596), complement(39256..39300),
complement(25552..25715), complement(24379..24480),
complement(20853..21103),
complement(AL355374.14:81450..81619),
complement(AL355374.14:80319..80435),
complement(AL355374.14:50238..50406),
complement(AL355374.14:44092..44401),
complement(AL355374.14:40790..41349),
complement(AL355374.14:37479..37532),
complement(AL355374.14:34005..34121),
complement(AL355374.14:30367..30744),
complement(AL355374.14:23606..27721))
/gene="ITIHS"
/locus_tag="RP11-676E7.1-001"
/product="inter-alpha (globulin) inhibitor H5"
/notes="match: ESTs: AAL34750.1 AA367857.1 AA367892.1
A1089902.1 A1090076.1 A1339610.1 A1417871.1 A14079707.1
AL540791.1 AL543337.1 AL544674.1 AL544702.1 AL550125.1
AL550297.1 AL550915.1 AL554626.1 AL5548595.1 AL5569892.1
AL5570737.1 AL574092.1 AL574207.1 AL574544.1 AL575373.1
AL577265.1 AW473748.1 BE893152.1 BE896223.1 BE898209.1
BF677721.1 BF689672.1 BF690566.1 BF726069.1 BF998687.1
BG006306.1 BG006608.1 BG282300.1 BG282310.1 BG286634.1
BG422562.1 BG435601.1 BG478666.1 BG676410.1 BG701677.1
BG767953.1 B104702.1 B1553925.1 B1559648.1 BM016522.1
BM542130.1 BM542378.1 BM549311.1 BM557411.1 BM675641.1
BM675641.1
complement(50428..50596), complement(39256..39300),
complement(25552..25715), complement(24379..24480),
complement(20853..21103),
complement(AL355374.14:81450..81619),
complement(AL355374.14:80319..80435),
complement(AL355374.14:50238..50406),
complement(AL355374.14:44092..44401),
complement(AL355374.14:40790..41349),
complement(AL355374.14:37479..37532),
complement(AL355374.14:34005..34121),
complement(AL355374.14:30367..30744),
complement(AL355374.14:23606..27721))
/gene="ITIHS"
/locus_tag="RP11-676E7.1-001"
/product="inter-alpha (globulin) inhibitor H5"
/notes="match: ESTs: AAL34750.1 AA367857.1 AA367892.1
A1089902.1 A1090076.1 A1339610.1 A1417871.1 A14079707.1
AL540791.1 AL543337.1 AL544674.1 AL544702.1 AL550125.1
AL550297.1 AL550915.1 AL554626.1 AL5548595.1 AL5569892.1
AL5570737.1 AL574092.1 AL574207.1 AL574544.1 AL575373.1
AL577265.1 AW473748.1 BE893152.1 BE896223.1 BE898209.1
BF677721.1 BF689672.1 BF690566.1 BF726069.1 BF998687.1
BG006306.1 BG006608.1 BG282300.1 BG282310.1 BG286634.1
BG422562.1 BG435601.1 BG478666.1 BG676410.1 BG701677.1
BG767953.1 B104702.1 B1553925.1 B1559648.1 BM016522.1
BM542130.1 BM542378.1 BM549311.1 BM557411.1 BM675641.1
BM675641.1
complement(50428..50596), complement(39256..39300),
complement(25552..25715), complement(24379..24480),
complement(20853..21103),
complement(AL355374.14:81450..81619),
complement(AL355374.14:80319..80435),
complement(AL355374.14:50238..50406),
complement(AL355374.14:44092..44401),
complement(AL355374.14:40790..41349),
complement(AL355374.14:37479..37532),
complement(AL355374.14:34005..34121),
complement(AL355374.14:30367..30744),
complement(AL355374.14:23606..27721))
/gene="ITIHS"
/locus_tag="RP11-676E7.1-001"
/product="inter-alpha (globulin) inhibitor H5"
/notes="match: ESTs: AAL34750.1 AA367857.1 AA367892.1
A1089902.1 A1090076.1 A1339610.1 A1417871.1 A14079707.1
AL540791.1 AL543337.1 AL544674.1 AL544702.1 AL550125.1
AL550297.1 AL550915.1 AL554626.1 AL5548595.1 AL5569892.1
AL5570737.1 AL574092.1 AL574207.1 AL574544.1 AL575373.1
AL577265.1 AW473748.1 BE893152.1 BE896223.1 BE898209.1
BF677721.1 BF689672.1 BF690566.1 BF726069.1 BF998687.1
BG006306.1 BG006608.1 BG282300.1 BG282310.1 BG286634.1
BG422562.1 BG435601.1 BG478666.1 BG676410.1 BG701677.1
BG767953.1 B104702.1 B1553925.1 B1559648.1 BM016522.1
BM542130.1 BM542378.1 BM549311.1 BM557411.1 BM675641.1
BM675641.1
complement(50428..50596), complement(39256..39300),
complement(25552..25715), complement(24379..24480),
complement(20853..21103),
complement(AL355374.14:81450..81619),
complement(AL355374.14:80319..80435),
complement(AL355374.14:50238..50406),
complement(AL355374.14:44092..44401),
complement(AL355374.14:40790..41349),
complement(AL355374.14:37479..37532),
complement(AL355374.14:34005..34121),
complement(AL355374.14:30367..30744),
complement(AL355374.14:23606..2
```



```

BM906016.1 BM931606.1 BM989082.1 BM993373.1 BM011450.1
BQ18379.1 BQ214877.1 BQ425746.1 BQ574118.1 BQ722286.1
BQ80658.1 BU165852.1 BU171456.1 BU189399.1 CA307415.1
CA414285.1 H51540.1 N26400.1 N26407.1 R88167.1 T48709.1
W22615.1
match: cDNAs: AB075833.1 AK027375.1 AK075381.1 AL833303.1"
join(complement(50428..50517),complement(39256..39300),
complement(25552..25715),complement(24379..24480),
complement(20853..21103),
complement(AL355374.14:81450..81619),
complement(AL355374.14:80319..80435),
complement(AL355374.14:50238..50406),
complement(AL355374.14:44092..44401),
complement(AL355374.14:40790..41349),
complement(AL355374.14:37479..37532),
complement(AL355374.14:34005..34121),
complement(AL355374.14:30367..30744),
complement(AL355374.14:27378..27721))
/genes="ITIH5"
/locus_tag="RP11-67687.1-001"
/standard_name="OTTHUMP00000019062"
/notes="match: proteins: Q8TF48 Q96K70"
/codon_start=1
/product="inter-alpha (globulin) inhibitor H5"
/protein_id="CAI12955.1"
/db_xref="GI:55958060"
/db_xref="GOA:Q5T665"
/db_xref="InterPro:IPR002035"
/db_xref="InterPro:IPR006587"
/db_xref="InterPro:IPR010600"
/db_xref="UniProt/TREMBL:Q5T665"
/translation="MLLLGLGLSLCVSGSEEAQSGHSEODGLRVPQVPELLOR
LKTPLMTFESVKTSILSYRAFTTVSCRLNRASEDQDIEFQMQIPAAFIINTMLI
GDKVYQGEITREKSGDRVKEKNGTTEENGKEGTIFRASAVIPSKDKAAFLSYE
ELLQRRCLKYRHSISVRPQQLSGRLSDVNLISAGIASLEVLPHNSRQSGRGE
DGGPPSTVINQNETFANIIEKPTVVOQARQAQNGILGDFIIRYDVNREQSIGDIQVL
NGYFVHPAPDKPLPKNVVFLDSSASVGTKLROTKDALFTILHDLRPQDRFSII
GFSNRKIKWKDHLISVTPDSIRDGKVIYHMSPTGGTDINGALQARILRLNKYVAHS
IGDSVSLVITDCKPTGTHIKILNNTREARGQVCIFTIGINDVDVFLLEKL
SLENGCTRRVHEEDAGSLIGFYDEIRLPLSDIRIDYPPSVVQATKTLFNYFN
GSEIIIAKLVDRKLDHLHVEVTASNSKKFIILKTDVPRPQAKGVTPGPGDG
EGDTNHIIRLWSYLTLLKELLSSQLQSDDEPEKRLRQQAALAVSYRFLPTSMKLR
GPVPRMGLSEAHGMSRAMPPEPVQSVRGAGTQGPLLKKPYQPRIKISYVDGDP
HFVDFPLSLRITVCENIDGQPGDILRLVSDHRDGSVTYNGELI GAPAPPNGHKQRTY
Query Match 100.0%; Score 22; DB 8; Length 174361;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAAGTGATCGCTGCCGTGGT 22
|||||
Db 171584 AGAAAGTGATCGCTGCCGTGGT 171563
|||||

RESULT 8
AL391687/c AL391687 194563 bp DNA linear HTG 20-OCT-2001
LOCUS Homo sapiens chromosome 10 clone RP11-299D3, 2 unordered pieces.
ACCESSION AL391687
VERSION AL391687.8 GI:14715355
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
REFERENCE 1
Almeida, J.
Direct Submission
Submitted (16-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
request: clonerequest@sanger.ac.uk
COMMENT On Jul 12, 2001 this sequence version replaced gi:13990116.

```

```

----- Sanger Centre
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA299D3
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 194262 bases at least Q40
Consensus quality: 194328 bases at least Q30
Consensus quality: 194372 bases at least Q20
Insert size: 194463; sum-of-contigs
Insert size: 173753; 41.5% error; agarose-fp
Quality coverage: 13.33x in Q20 bases; sum-of-contigs Quality
coverage: 14.92x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 172067: contig of 172067 bp in length
* 172068 172167: gap of 100 bp
* 172168 194563: contig of 22396 bp in length.
FEATURES             Location/Qualifiers
     source            1..194563
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-299D3"
                     /clone_lib="RPCI-11.2"
     misc_feature       1..172067
                     /note="assembly fragment:00152
                     fragment_chain:1"
     misc_feature       172168..194563
                     /note="assembly fragment:03892
                     fragment_chain:1
                     clone_end:T7
                     vector_side:right"
ORIGIN
Query Match 100.0%; Score 22; DB 14; Length 194563;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAAGTGATCGCTGCCGTGGT 22
|||||
Db 90730 AGAAAGTGATCGCTGCCGTGGT 90709
|||||

RESULT 9
AC017711/c AC017711 7111 bp DNA linear HTG 10-DEC-1999
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
DEFINITION AC017711
ACCESSION AC017711
VERSION AC017711.1 GI:6554286
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 7111)
AUTHORS Adams, M. and Venter, J.C.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-1999) Celera Genomics, 45 West Gude Drive,

```

Rockville, MD, USA
 This sequence was identified as CDM:10211824 by the submitter.
 For more information on this record e-mail to fly@celera.com.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES

source
 1. 7111
 /location/Qualifiers
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"

ORIGIN

Query Match 83.6%; Score 18.4; DB 14; Length 7111;
 Best Local Similarity 95.0%; Pred. No. 7.2e+02;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GAAAGTATCGTCCGCTGG 21

Db 6348 GAAAGTATCGCTACCGTGG 6329

RESULT 10

CR847851 166348 bp DNA linear HTG 15-JUN-2005
 LOCUS
 DEFINITION Danio rerio chromosome 14 clone DKEYP-115D7, WORKING DRAFT
 SEQUENCE, 10 unordered pieces.

ACCESSION CR847851

VERSION CR847851.4 GI:67845775

KEYWORDS HTG; HTGS; PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE Danio rerio (zebrafish)

ORGANISM

Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.

REFERENCE

AUTHORS McLaren, S.

TITLE Direct Submission

JOURNAL Submitted (14-JUN-2005) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 zfish-help@sanger.ac.uk Clone requests:

http://www.sanger.ac.uk/Projects/D/rerio/faqs.shtml#dataeight
 On Jun 15, 2005 this sequence version replaced gi:54019839.

COMMENT

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: zfish-help@sanger.ac.uk

----- Project Information

Center project name: zKp115D7

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator; 100% of reads

Consensus quality: 161693 bases at least Q40

Consensus quality: 162463 bases at least Q30

Consensus quality: 163159 bases at least Q20

Insert size: 165448; sum-of-contigs

Insert size: 191089; 11.8% error; agarose-fp

Quality coverage: 8.42x in Q20 bases; sum-of-contigs Quality
 coverage: 7.39x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 10 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 4408: contig of 4408 bp in length

* 4409 4508: gap of 100 bp

* 4509 15390: contig of 10882 bp in length

* 15391 15490: gap of 100 bp
 * 15491 30342: contig of 14852 bp in length
 * 30343 30442: gap of 100 bp
 * 30443 77547: contig of 47105 bp in length
 * 77548 77647: gap of 100 bp
 * 77648 113849: contig of 36202 bp in length
 * 113850 113949: gap of 100 bp
 * 113950 119804: contig of 5855 bp in length
 * 119805 119904: gap of 100 bp
 * 119905 124866: contig of 4962 bp in length
 * 124867 124966: gap of 100 bp
 * 124967 135191: contig of 10225 bp in length
 * 135192 135291: gap of 100 bp
 * 135292 149612: contig of 14321 bp in length
 * 149613 149712: gap of 100 bp
 * 149713 166348: contig of 16636 bp in length.

FEATURES

source

1. 166348
 /location/Qualifiers
 /organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /chromosome="14"
 /clone="DKEYP-115D7"
 /clone_lib="DanioKeyPilot"

misc_feature

1. 4408
 /notes="assembly_fragment:00030
 fragment_chain:1"
 4509. 15390
 /notes="assembly_fragment:00274
 fragment_chain:1"

misc_feature

15491. 30342
 /notes="assembly_fragment:00615
 fragment_chain:1"
 30443. 77547
 /notes="assembly_fragment:01921
 fragment_chain:2"

misc_feature

77648. 113849
 /notes="assembly_fragment:01377
 fragment_chain:2"
 113950. 119804
 /notes="assembly_fragment:00183
 fragment_chain:2"

misc_feature

119905. 124866
 /notes="assembly_fragment:00099.0"
 124967. 135191
 /notes="assembly_fragment:00425"
 135292. 149612
 /notes="assembly_fragment:00835"

misc_feature

149713. 166348
 /notes="assembly_fragment:01070"

misc_feature

ORIGIN

Query Match 83.6%; Score 18.4; DB 14; Length 166348;

Best Local Similarity 95.0%; Pred. No. 6e+02;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGAAAGTATCGTCCGCTGG 20

Db 97301 AGAAAGTATCGTCCGCTGG 97320

RESULT 11

AC146480/c

LOCUS

DEFINITION Danio rerio clone CH211-4009, complete sequence.

ACCESSION AC146480

VERSION AC146480.2 GI:37951362

KEYWORDS HTG.

SOURCE Danio rerio (zebrafish)

ORGANISM

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Danio.

VERT 24-OCT-2003

1 (bases 1 to 166906)
Talbot,W.S., Rauch,G.J., Noonan,J., Grimwood,J., Dickson,M., Schmutz,J. and Myers,R.M.
Genomic studies of vertebrate diversity
Unpublished
REFERENCE
AUTHORS
2 (bases 1 to 166906)
Talbot,W.S., Rauch,G.J., Noonan,J., Grimwood,J., Dickson,M., Schmutz,J. and Myers,R.M.
Direct Submission
Submitted (19-AUG-2003) Stanford Human Genome Center, 975 California Avenue, Palo Alto, CA 94304, USA
REFERENCE
AUTHORS
3 (bases 1 to 166906)
Talbot,W.S., Rauch,G.J., Noonan,J., Grimwood,J., Dickson,M., Schmutz,J. and Myers,R.M.
Direct Submission
Submitted (24-OCT-2003) Stanford Human Genome Center, 975 California Avenue, Palo Alto, CA 94304, USA
TITLE
JOURNAL
COMMENT
On Oct 24, 2003 this sequence version replaced gi:33859862.
The sequence of the clone was established as a mapping and sequencing collaboration at the Stanford Genome Evolution Center, funded by the NIH Centers of Excellence in Genomic Science (CEGS) initiative (<http://cegs.stanford.edu>). The clone was isolated from the BAC library CHORI-211 (<http://bacpac.chori.org>).
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
FEATURES
source
1..166906
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-4009"
ORIGIN
Query Match 83.6%; Score 18.4; DB 5; Length 166906;
Best Local Similarity 95.0%; Pred. No. 6e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 AGAAAGTCATCGCTGCCGTG 20
Db 146447 AGAAAGTCATCGCTGCCATG 146428
RESULT 12
AL929558/c
LOCUS
DEFINITION
Zebrafish DNA sequence from clone CH211-159N7, complete sequence.
ACCESSION
AL929558
VERSION
AL929558.13 GI:38143466
KEYWORDS
HTG.
SOURCE
Danio rerio (zebrafish)
ORGANISM
Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE
1 (bases 1 to 170597)
Whitehead,S.
Direct Submission
Submitted (01-NOV-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Nov 1, 2003 this sequence version replaced gi:37718665.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: zfish-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.
This sequence was finished as follows otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: En: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep. Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.
Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhiyong Bao and Sean Eddy, submitted), and those beginning 'drr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml CH211-159N7 is from a CHORI-211 BAC library
VECTOR: pTARBAC2.1.
FEATURES
Location/Qualifiers
source
1..170597
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-159N7"
/clone_lib="CHORI-211"
ORIGIN
Query Match 83.6%; Score 18.4; DB 5; Length 170597;
Best Local Similarity 95.0%; Pred. No. 5.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 AGAAAGTCATCGCTGCCGTG 20
Db 146404 AGAAAGTCATCGCTGCCATG 146385
RESULT 13
AC018491
LOCUS
DEFINITION
Drosophila melanogaster clone BACR11H15, complete sequence.
ACCESSION
AC018491
VERSION
AC018491.9 GI:50872279
KEYWORDS
HTG.
SOURCE
Drosophila melanogaster (fruit fly)
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 177953)
Celisner,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D., Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
Direct Submission
Submitted (13-DEC-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
2 (bases 1 to 177953)
Celisner,S., Carlson,J., Wan,K., Pfeiffer,B., Frise,E., George,R., Hoskins,R., Stapleton,M., Pacleb,J., Park,S., Svirskas,R., Smith,E., Yu,C. and Rubin,G.
Direct Submission
Submitted (31-JUL-2004) Berkeley Drosophila Genome Project, MS

64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, US

On Jul 31, 2004 this sequence version replaced gi:14701992.

Sequence submitted by:

Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory
Berkeley, CA 94720

This sequence submission incorporates changes made during reevaluation of the assembly or fingerprint verification of the clone. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to bdg@fruitfly.org.

FEATURES

source

```
1. 177953
   Location/Qualifiers
     /organism="Drosophila melanogaster"
     /mol_type="genomic DNA"
     /strain="y; cn bw sp"
     /db_xref="taxon:7227"
     /chromosome="X"
     /map="19A-198"
     /clone="BAC clone BACR11H15 (D1225)"
     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBACs.6)"
```

ORIGIN

```
Query Match      83.6%; Score 18.4; DB 2; Length 177953;
Best Local Similarity 95.08; Pred. No. 5.9e+05;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

Oy 2 GAAAGTGATCGTGCCTGG 21

```
|||||
Db 146451 GAAAGTGATCGTACCCTGG 146470
```

RESULT 14

AE002611

LOCUS

AB0002611 297848 bp DNA linear INV 10-AUG-2005
Drosophila melanogaster chromosome X, section 66 of 74 of the
complete sequence.

ACCESSION AE002611 AE014298

VERSION AE002611.5 GI:55380356

KEYWORDS

SOURCE Drosophila melanogaster (fruit fly)

ORGANISM

Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 297848)

Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Ananidis,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazer,R.G., Champe,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor,G.L.,
Abril,J.F., Agbayani,A., An,H.J., Andrews-Pfannkoch,C., Baldwin,D.,
Ballew,R.M., Basu,A., Baxendale,J., Bayraktaroglu,L., Beasley,E.M.,
Beeson,K.Y., Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S.,
Borkova,D., Botchan,M.R., Bouck,J., Brokstein,P., Brotter,P.,
Burtis,K.C., Busam,D.A., Butler,H., Cadieu,E., Center,A.,
Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B.,
Davies,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I.,
Dietz,S.M., Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S.,
Dunkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C.,
Ferriera,S., Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
Geibart,M.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Laeko,P., Lei,Y., Levisky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,

McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nuskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
Saunders,R.D., Scheeler,P., Shen,H., Shue,B.C., Siden-Kiamos,I.,
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Svirekas,R., Rector,C., Turner,R.,
Venter,B., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)

10731132

JOURNAL
PUBMED
AUTHORS

2 (bases 1 to 297848)

Celniker,S.E., Wheeler,D.A., Kronmiller,B., Carlson,J.W.,
Halpern,A., Patel,S., Adams,M., Champe,M., Dugan,S.P., Frise,E.,
Hodgson,A., George,R.A., Hoskins,R.A., Laverty,T., Muzny,D.M.,
Nelson,C.R., Pacleb,J.M., Park,S., Pfeiffer,B.D., Richards,S.,
Sodergren,E.J., Svirekas,R., Tabor,P.E., Wan,K., Stapleton,M.,
Sutton,G.G., Venter,C., Weinstock,G., Scherer,S.E., Myers,E.W.,
Gibbs,R.A. and Rubin,G.M.
Finishing a whole-genome shotgun: release 3 of the Drosophila
melanogaster euchromatic genome sequence
Genome Biol. 3 (12), RESEARCH0079 (2002)
12537568

3 (bases 1 to 297848)

JOURNAL
PUBMED
AUTHORS

Misra,S., Crosby,M.A., Mungall,C.J., Matthews,B.B., Campbell,K.S.,
Hradecky,P., Huang,Y., Kaminker,J.S., Milburn,G.H., Prochnik,S.E.,
Smith,C.D., Tupy,J.L., Whitfield,E.J., Bayraktaroglu,L.,
Berman,B.P., Bettencourt,B.R., Celniker,S.E., de Grey,A.D.,
Drysdale,R.A., Harris,N.L., Richter,J., Russo,S., Schroeder,A.J.,
Shu,S.Q., Stapleton,M., Yamada,C., Ashburner,M., Gelbart,W.M.,
Rubin,G.M. and Lewis,S.E.
Annotation of the Drosophila melanogaster euchromatic genome: a
systematic review
Genome Biol. 3 (12), RESEARCH0083 (2002)
12537572

4 (bases 1 to 297848)

JOURNAL
PUBMED
AUTHORS

Kaminker,J.S., Bergman,C.M., Kronmiller,B., Carlson,J.,
Svirekas,R., Patel,S., Frise,E., Wheeler,D.A., Lewis,S.E.,
Rubin,G.M., Ashburner,M. and Celniker,S.E.
The transposable elements of the Drosophila melanogaster
euchromatin: a genomics perspective
Genome Biol. 3 (12), RESEARCH0084 (2002)
12537573

5 (bases 1 to 297848)

JOURNAL
PUBMED
AUTHORS

Celniker,S., Carlson,J., Wan,K., Pfeiffer,B., Frise,E., George,R.,
Hoskins,R., Stapleton,M., Pacleb,J., Park,S., Svirekas,R.,
Smith,E., Yu,C. and Rubin,G.

Berkeley Drosophila Genome Project

Drosophila melanogaster release 4 sequence

Unpublished

6 (bases 1 to 297848)

JOURNAL
PUBMED
AUTHORS

Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
7 (bases 1 to 297848)

8 (bases 1 to 297848)

JOURNAL
PUBMED
AUTHORS

FlyBase
Direct Submission
Submitted (22-JUL-2005) FlyBase, Harvard University, Biological
Laboratories, 16 Divinity Ave, Cambridge, MA 02138, USA
On Nov 4, 2004 this sequence version replaced gi:28381542.
Location/Qualifiers

```
1. 297848
   /organism="Drosophila melanogaster"
   /mol_type="genomic DNA"
   /db_xref="taxon:7227"
```


Query Match 80.9%; Score 17.8; DB 1; Length 460;
Best Local Similarity 90.5%; Pred.No. 1.6e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GAAAGTGATCGCTGCCGTGGT 22
|||
Db 131 GACAGTGATCGCTGCCGTGGT 151
|||

Search completed: November 26, 2005, 19:11:04
Job time : 121.932 secs

KEYWORDS
STS

SOURCE
Macaca mulatta (rhesus monkey)

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Cercopithecoidea; Cercopithecoidea; Macaca.

REFERENCE
1 (bases 1 to 616)
Spindel, B.R., Pauley, M., Jia, Y., Thompson, S., Lanthorn, M., Gravett, C., Lupo, S.L., Tchourbanov, A., Ali, H., Ojeda, S.R., and Norgren, R. B.

AUTHORS
Targeted amplification of the 3' end of rhesus macaque orthologs of human genes
Unpublished (2004)

JOURNAL
Contact: Spindel, B.R.
Division of Neuroscience
Oregon National Primate Research Center
505 NW 185th Avenue, Beaverton, OR 97006, USA
Tel: 403-690-5388
Fax: 503-690-5384
Email: spindel@ohsu.edu
Primer A: ggaattcattatgaagcatt
Primer B: tacattagtcgtctcagc
STS size: 616

COMMENT
PCR Profile:
Hot Start: 95 degrees C for 2.00 min
Denaturation: 95 degrees C for 0.50 min
Annealing: 51 degrees C for 0.50 min
Polymerization: 72 degrees C for 1.00 min
PCR Cycles: 35
Extension: 72 degrees C for 7.0 min
Thermal Cycler: MJ Instruments PTC100

Protocol:
Template: 200 ng
Primer: each 100 μM
dNTP's: each 200 μM
Tag Polymerase: 0.05 units/μl (Fast Start High Fidelity, Roche)
Total Vol: 50 μl

Buffer:
MgCl2: 1.8 mM
Fast Start polymerase reaction buffer (Roche)

FEATURES
source
1..616
location/Qualifiers
/organism="Macaca mulatta"
/mol_type="genomic DNA"
/strain="Indian origin"
/db_xref="taxon:9544"
/clone="MMA9291"
/clone_1b="Rhesus macaque genomic DNA"
/dev_stage="Adult"
/note="Organ: Liver; Vector: pGEM-T Easy; V-type: Plasmid; STS was amplified from rhesus genomic DNA with the human forward and reverse primers listed above and subcloned into pGEM-T Easy"
1..616
/gene="KIN"
/note="KIN, antigenic determinant of rca protein homolog (mouse)"
STS
1..616

ORIGIN
/gene="KIN"

Query Match
Best Local Similarity 100.0%; Score 26; DB 10; Length 616;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OR
1 GCGACACCAATTTGATCTTTAGA 26
71 GCGACACCAATTTGATCTTTAGA 46

Db
1002 GCGACACCAATTTGATCTTTAGA 977

RESULT 3
AX003310/c 1002 bp DNA linear PAT 24-AUG-2000
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo;

REFERENCE
AUTHORS
TITLE
JOURNAL
ANGULO-MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE (FR)

FEATURES
source
1..1002
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match
Best Local Similarity 100.0%; Score 26; DB 6; Length 1002;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OR
1 GCGACACCAATTTGATCTTTAGA 26
1002 GCGACACCAATTTGATCTTTAGA 977

RESULT 4
AX003308/c 1528 bp DNA linear PAT 24-AUG-2000
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo;

REFERENCE
AUTHORS
TITLE
JOURNAL
ANGULO-MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE (FR)

FEATURES
source
1..1528
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match
Best Local Similarity 100.0%; Score 26; DB 6; Length 1528;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Wed Nov 30 09:40:42 2005

us-09-555-529-19.rge

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GCGAACACCAATTGATGCTTTAAGA 26
1303 GCGAACACCAATTGATGCTTTAAGA 1278

RESULT 5
HSAJ5273/c
LOCUS
DEFINITION Homo sapiens mRNA for Kin17 protein.
ACCESSION AJ005273
VERSION AJ005273.1 GI:3850703
KEYWORDS KIN17 gene.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

REFERENCE
AUTHORS Kannouche, P., Mauffrey, P., Pinon-Lataillade, G., Mattei, M.G.,
Sarasin, A., Daya-Grosjean, L. and Angulo, J.F.
TITLE Molecular cloning and characterization of the human KIN17 cDNA
encoding a component of the UVC response that is conserved among
metazoans
JOURNAL Carcinogenesis 21 (9), 1701-1710 (2000)
PUBMED 10964102

REFERENCE
AUTHORS Mauffrey, P.
TITLE Direct Submission
JOURNAL Submitted (08-APR-1998) Mauffrey P., LGR, CEA-DSV-DRR, BP 6 -
Fontenay aux Roses Cedex, 92265, FRANCE
REMARK Revised by [3]
AUTHORS Mauffrey, P.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1998) Mauffrey P., LGR, CEA-DSV-DRR, BP 6 -
Fontenay aux Roses Cedex, 92265, FRANCE
COMMENT On Nov 7, 1998 this sequence version replaced gi:3046739.
Kannouche P., Mauffrey P., Pinon-Lataillade G., Biard D.S.F.,
Angulo J.F.
Sequences codant pour la Proteine kin17 et leurs Applications -
French Patent Nr 97 15536 1:1-50(1997).

FEATURES
source
1..1528
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/sex="male"
/tissue_type="testis"
1..1528
/gene="Kin17"
66..1247
/gene="Kin17"
/codon_start=1
/protein_id="CAA06462.1"
/db_xref="GI:3850704"
/db_xref="GOA:O60870"
/db_xref="UniProt/TrEMBL:O60870"
/translation="MGKSPDLTPKAIANRIKSGLOKLRWYCOMCKOCRDENGPKCH
CMSEHORALLASNPQPMDFSEFRNDFLELRPRGTRVHNNIVVYISHR
EHIDMQETUETDTTKKRGGLCKVDPTPKGWYIQYIDRDPETIPKOLEKKK
QDDDEKTAKEIEQVRGLGKQEVPTIELSRENDSEKVTIPLNSKAGSSGAT
SSKSTTSPALKTIGSSAVKREKSSQSTQSKKKKKKALDEIMEIEEKKTAR
TDYLOPEIIVK1ITKKGEKVKKAIKVEIDKYTAVVKMIDSGDKLKLQTHLET
VIPAGKRILVNGGRVNEGTLSEINEKFTSATIVTETPLKGRVEGIQYEDISK
A"

gene
CDS

Query Match 100.0%; Score 26; DB 8; Length 1528;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCGAACACCAATTGATGCTTTAAGA 26
Db 1303 GCGAACACCAATTGATGCTTTAAGA 1278

RESULT 6
BC017309/c
LOCUS
DEFINITION BC017309 2069 bp mRNA linear PRI 29-JUN-2004
Homo sapiens KIN, antigenic determinant of recA protein homolog
(mouse). mRNA (CDNA clone MGC:29595 IMAGE:5089578), complete cds.
ACCESSION BC017309
VERSION BC017309.2 GI:40226034
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

REFERENCE
AUTHORS Strausberg, R.L., Feingold, S.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, P.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buettow, K.H., Schaefer, C.P., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P.,
Diatchenko, L., Marusik, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Scapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullen, S.J., Bosak, S.A., McSwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shvachenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, B.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywicki, M.I., Skalek, U., Smal, U.,
Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
TITLE Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED 12477932

REFERENCE
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT On Dec 19, 2003 this sequence version replaced gi:16878214.
Contact: MGC help desk
Email: cgabs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisec.nih.gov/>
Contact: nisc@nih.gov
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Districh, N.L., Granice, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-t., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantrick, S., Thomas, P.J., Touchman, J.W.,
Tauragon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, D.,
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IMAGE Plate: 39 Row: 0 Column: 18
This clone was selected for full length sequencing because it

FEATURES
Source

passed the following selection criteria: matched mRNA gi: 13124882.

Location/Qualifiers

1. 2069
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:29595 IMAGR:5089578"
/clone_type="Pancras", epithelioid carcinoma"
/clone_id="NIH_MGC_42"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
1. 2069
/gene="KIN"
/note="synonym: KIN17, BTCD"
/db_xref="GeneID:22944"
/db_xref="MIM:601720"
12. 1193
/gene="KIN"
/codon_start=1
/product="Hsk17 protein"
/protein_id="AAH17309.1"
/db_xref="GI:16878215"
/db_xref="GeneID:22944"
/db_xref="MIM:601720"
/translation="MGKSDPLTPKAIANKSKGQKLRWYCOMCKQCDENGFKCH
CHSEHROQLASENPOQPMDFSEFRNDLRLRRPOTKRNHNTVINYETLSHR
EHIHNATQWETLTDPTKMGREGCLCKVDPTKGYIYIDRPTIRQLLEKKK
QDLDEKTAKEIIEQVRGIEGKEQVPTFELRENDERKTFNLSKACSSCAT
SSKSTSLPSALKTIGSSAKRKSOSSTOSKSKKKKSKALDEIIEEKKRAT
TVMQAPRIIVKTIITKLGKHKKAIKVEYIDVTAVVADISGDKLDDTHLET
VIPAGKRILVNGVGRNGEGLSINEKTFATVIVETGPLKGRVVEGIQYEDISK
A"

gene

CDS

ORIGIN

Query Match 100.0%; Score 26; DB 8; Length 2069;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCGACACCAATTGATGCTTTAAGA 26
Db 1249 GCGACACCAATTGATGCTTTAAGA 1224

RESULT 7
AL158044
LOCUS 174361 bp DNA linear PRI 18-MAY-2005
DEFINITION Human DNA sequence from clone RP11-264C14 on chromosome 10 contains
the 5' end of the gene for a novel protein (MGC10848), the ITH2
gene for inter-alpha (globulin) inhibitor, H2 polypeptide (H2P),
the KIN gene for KIN, antigenic determinant of reca protein homolog
(mouse), the 3' end of the ATP5C1 gene for ATP synthase, H4
transporting, mitochondrial F1 complex, gamma polypeptide 1,
(ATP5C1, ATP5C1), a novel gene and two Cpg islands, complete
sequence.

AL158044
VERSION AL158044.17 GI:16214543
KEYWORDS HMG; ATP5C1; ATP5C1; Cpg island; H2P; ITH2; KIN; MGC10848.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo
1 (bases 1 to 174361)

REFERENCE 1 (bases 1 to 174361)

AUTHORS Direct Submissions
TITLE Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgehire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
On Oct 17, 2001 this sequence version replaced gi:15131961.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL, Sw: SWISSPROT, Tr: TREMBL, Wp: WORMPEP, information
on the WORMPEP database can be found at

FEATURES
Source

http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
RP11-264C14 is from the library RP11-11.1 constructed by the group
of Peter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vegas@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one subclone; and the assembly was confirmed by restriction digest,
except on the rare occasion of the clone being a YAC.

1. 174361
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-264C14"
/clone_1lb="RP11-11.1"

misc_feature
gene

1
/note="Clone left end: RP11-264C14"
join(complement(13206..3285),
complement(AL355374.14:26760..81619))
/gene="ITH5"

1
/locus_tag="RP11-676E7.1-003"
join(complement(13206..3285),
complement(AL355374.14:81450..81619),
complement(AL355374.14:80319..80435),
complement(AL355374.14:50238..50406),
complement(AL355374.14:44092..44401),
complement(AL355374.14:40790..41349),
complement(AL355374.14:37479..37532),
complement(AL355374.14:34005..34121),
complement(AL355374.14:30367..30744),
complement(AL355374.14:26760..27721))
/gene="ITH5"

1
/locus_tag="RP11-676E7.1-003"
/product="inter-alpha (globulin) inhibitor H5"
/note="match: ESTs: BE386984.1 BB387282.1 BG333950.1
BG466698.1 B0225649.1
match: CDNA: BC004282.1"
join(complement(13206..3215),
complement(AL355374.14:81450..81619),
complement(AL355374.14:80319..80435),
complement(AL355374.14:50238..50406),
complement(AL355374.14:44092..44401),
complement(AL355374.14:40790..41349),
complement(AL355374.14:37479..37532),
complement(AL355374.14:34005..34121),
complement(AL355374.14:30367..30744),
complement(AL355374.14:27378..27721))
/gene="ITH5"

1
/locus_tag="RP11-676E7.1-003"
/standard_name="OTTHOMP0000019064"
/note="match: proteins: Q9B979"
/codon_start=1
/product="inter-alpha (globulin) inhibitor H5"
/protein_id="CAI12953.1"
/db_xref="GI:55958058"
/db_xref="GCA:Q5T664"
/db_xref="InterPro:IPR002035"
/db_xref="InterPro:IPR010600"

CDS

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:41:08 ; Search time 7.76471 Seconds
(without alignments)
421.056 Million cell updates/sec

Title: US-09-555-529-18

Perfect score: 22

Sequence: 1 agaaagtgcgtcgctgggt 22

Scoring table: IDENTITY|NJC

Gapop 10|0 , Gapext 1.0

Searched: 3205263 seqs, 74304013 residues

Total number of hits satisfying chosen parameters: 6410526

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA New:*

1: /cgn2_6/ptodata/2/pubna/US10_NEW_PUB.seq.*
2: /cgn2_6/ptodata/2/pubna/US06_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubna/US07_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubna/US08_NEW_PUB.seq.*
5: /cgn2_6/ptodata/2/pubna/US09_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubna/PCT_NEW_PUB.seq.*
7: /cgn2_6/ptodata/2/pubna/US11_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubna/US11_NEW_PUB.seq.*
9: /cgn2_6/ptodata/2/pubna/US11_NEW_PUB.seq.*
10: /cgn2_6/ptodata/2/pubna/US60_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	15.2	69.1	333	1	US-10-802-796-723 Sequence 723, App
C 2	14.6	66.4	163	1	US-10-802-796-492 Sequence 492, App
C 3	14.6	66.4	262	1	US-10-802-796-275 Sequence 275, App
C 4	14.6	66.4	264	1	US-10-802-796-228 Sequence 228, App
C 5	14.6	66.4	295	1	US-10-802-796-448 Sequence 448, App
C 6	14.6	66.4	325	1	US-10-802-796-216 Sequence 216, App
C 7	14.6	66.4	420	1	US-10-802-796-496 Sequence 496, App
C 8	14.6	66.4	428	9	US-11-082-389-413 Sequence 413, App
C 9	14.6	66.4	444	1	US-10-821-234-418 Sequence 418, App
C 10	14.6	66.4	445	1	US-10-802-796-75 Sequence 75, Appl
C 11	14.6	66.4	470	1	US-10-802-796-157 Sequence 157, App
C 12	14.6	66.4	1575	1	US-10-821-234-779 Sequence 779, App
C 13	14.2	64.5	19	8	US-11-101-244-16138 Sequence 16138, A
C 14	14.2	64.5	19	9	US-11-083-784-16138 Sequence 16138, A
C 15	14.2	64.5	1018	1	US-10-821-234-225 Sequence 225, App
C 16	14	63.6	1277	1	US-10-821-234-9 Sequence 9, Appli
C 17	14	63.6	1511	7	US-11-102-978-6 Sequence 6, Appli
C 18	13.8	62.7	1509	9	US-11-082-389-411 Sequence 411, App
C 19	13.8	62.7	3341	7	US-11-110-837-1 Sequence 1, Appli
C 20	13.8	62.7	3341	7	US-11-110-837-3 Sequence 3, Appli
C 21	13.6	61.8	247	1	US-10-802-796-475 Sequence 475, App
C 22	13.6	61.8	889	9	US-11-082-389-119 Sequence 119, App
C 23	13.6	61.8	957	9	US-11-082-389-117 Sequence 117, App

Sequence 105, App
Sequence 201, App
Sequence 103, App
Sequence 57, Appl
Sequence 127, App
Sequence 233, App
Sequence 261, App
Sequence 213019,
Sequence 1021672,
Sequence 1035689,
Sequence 1031019,
Sequence 1021672,
Sequence 1035689,
Sequence 1035689,
Sequence 139395,
Sequence 452828,
Sequence 1250329,
Sequence 149736,
Sequence 1449781,
Sequence 139395,
Sequence 452828,
Sequence 1250329,
Sequence 1449736,

ALIGNMENTS

RESULT 1
US-10-802-796-723/c
; Sequence 723, Application US/10802796
; Publication No. US20050250104A1
; GENERAL INFORMATION:
; APPLICANT: COLE, STEWART
; APPLICANT: BUCHRIESER-BROSCH, ROLAND
; APPLICANT: GORDON, STEPHEN
; APPLICANT: BILLAULT, ALAIN
; TITLE OF INVENTION: A METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST
; TITLE OF INVENTION: FROM THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED
; TITLE OF INVENTION: DNA LIBRARY APPLICATION TO THE DETECTION OF
; TITLE OF INVENTION: MYCOBACTERIA.
; FILE REFERENCE: 05394.0011-00000
; CURRENT APPLICATION NUMBER: US/10/802,796
; PRIOR FILING DATE: 2004-03-18
; PRIOR APPLICATION NUMBER: US/09/673,476
; PRIOR FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: PCT/IB99/00740
; PRIOR FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 09/060,756
; PRIOR FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: PatentIn Ver. 2.2
; SEQ ID NO 723
; LENGTH: 333
; TYPE: DNA
; ORGANISM: Mycobacterium bovis
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (45)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (64)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (130)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (146)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base

```
; LOCATION: (205)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (211)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (271)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (309)
; OTHER INFORMATION: a, t, c or g
US-10-802-796-723
```

```
Query Match 69.1%; Score 15.2; DB 1; Length 333;
```

```
Best Local Similarity 81.0%; Pred. No. 57;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 2 GAAAGTATCGCTGCCGTGGT 22
||| ||| ||| ||| ||| ||| |||
Db 133 GAANGTGTGCTGGCGTGT 113
```

RESULT 2

```
US-10-802-796-492/c
; Sequence 492, Application US/10802796
; Publication No. US20050250104A1
; GENERAL INFORMATION:
; APPLICANT: COLE, STEWART
; APPLICANT: BUCHRIESER-BROSCH, ROLAND
; APPLICANT: GORDON, STEPHEN
; APPLICANT: BILLAULT, ALAIN
; TITLE OF INVENTION: A METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST
; TITLE OF INVENTION: FROM THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED
; TITLE OF INVENTION: DNA LIBRARY. APPLICATION TO THE DETECTION OF
; TITLE OF INVENTION: MYCOBACTERIA.
; FILE REFERENCE: 05394.0011-00000
; CURRENT APPLICATION NUMBER: US/10/802,796
; CURRENT FILING DATE: 2004-03-18
; PRIOR FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: PCT/IB99/00740
; PRIOR FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 09/060,756
; PRIOR FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: PatentIn Ver. 2.2
; SEQ ID NO 492
; LENGTH: 163
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-10-802-796-492
```

```
Query Match 66.4%; Score 14.6; DB 1; Length 163;
```

```
Best Local Similarity 81.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 2 GAAAGTATCGCTGCCGTGGT 22
||| ||| ||| ||| ||| ||| |||
Db 103 GAATGTGTGCTGGCGTGT 83
```

RESULT 3

```
US-10-802-796-275/c
; Sequence 275, Application US/10802796
; Publication No. US20050250104A1
; GENERAL INFORMATION:
; APPLICANT: COLE, STEWART
; APPLICANT: BUCHRIESER-BROSCH, ROLAND
; APPLICANT: GORDON, STEPHEN
; APPLICANT: BILLAULT, ALAIN
```

```
; TITLE OF INVENTION: A METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST
; TITLE OF INVENTION: FROM THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED
; TITLE OF INVENTION: DNA LIBRARY. APPLICATION TO THE DETECTION OF
; TITLE OF INVENTION: MYCOBACTERIA.
; FILE REFERENCE: 05394.0011-00000
; CURRENT APPLICATION NUMBER: US/10/802,796
; CURRENT FILING DATE: 2004-03-18
; PRIOR APPLICATION NUMBER: US/09/673,476
; PRIOR FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: PCT/IB99/00740
; PRIOR FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 09/060,756
; PRIOR FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: PatentIn Ver. 2.2
; SEQ ID NO 275
; LENGTH: 262
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (21)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (39)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (97)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (122)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (130)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (144)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (222)..(223)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (225)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (229)
; OTHER INFORMATION: a, t, c or g
US-10-802-796-275
```

```
Query Match 66.4%; Score 14.6; DB 1; Length 262;
```

```
Best Local Similarity 81.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 2 GAAAGTATCGCTGCCGTGGT 22
||| ||| ||| ||| ||| ||| |||
Db 66 GAATGTGTGCTGGCGTGT 46
```

RESULT 4

```
US-10-802-796-228/c
; Sequence 228, Application US/10802796
; Publication No. US20050250104A1
; GENERAL INFORMATION:
; APPLICANT: COLE, STEWART
; APPLICANT: BUCHRIESER-BROSCH, ROLAND
```

```
; APPLICANT: GORDON, STEPHEN
; APPLICANT: BILLAULT, ALAIN
; TITLE OF INVENTION: A METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST
; TITLE OF INVENTION: FROM THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED
; TITLE OF INVENTION: DNA LIBRARY. APPLICATION TO THE DETECTION OF
; TITLE OF INVENTION: MYCOBACTERIA.
; FILE REFERENCE: 05394.0011-00000
; CURRENT APPLICATION NUMBER: US/10/802,796
; CURRENT FILING DATE: 2004-03-18
; PRIOR APPLICATION NUMBER: US/09/673,476
; PRIOR FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: PCT/IB99/00740
; PRIOR FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 09/060,756
; PRIOR FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: PatentIn Ver. 2.2
; SEQ ID NO 228
; LENGTH: 264
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; US-10-802-796-228

Query Match      66.4%; Score 14.6; DB 1; Length 264;
Best Local Similarity 81.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy      2  GAAAGTGATCGTCGCGCTGGT 22
Db      132 GAATGTGGTCGCTGGCGTCGT 112

RESULT 5
US-10-802-796-448/c
; Sequence 448, Application US/10802796
; Publication No. US20050250104A1
; GENERAL INFORMATION:
; APPLICANT: COLE, STEWART
; APPLICANT: BUCHRIESER-BROSCH, ROLAND
; APPLICANT: GORDON, STEPHEN
; APPLICANT: BILLAULT, ALAIN
; TITLE OF INVENTION: A METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST
; TITLE OF INVENTION: FROM THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED
; TITLE OF INVENTION: DNA LIBRARY. APPLICATION TO THE DETECTION OF
; TITLE OF INVENTION: MYCOBACTERIA.
; FILE REFERENCE: 05394.0011-00000
; CURRENT APPLICATION NUMBER: US/10/802,796
; CURRENT FILING DATE: 2004-03-18
; PRIOR APPLICATION NUMBER: US/09/673,476
; PRIOR FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: PCT/IB99/00740
; PRIOR FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 09/060,756
; PRIOR FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: PatentIn Ver. 2.2
; SEQ ID NO 448
; LENGTH: 295
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (136)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (268)..(269)
; OTHER INFORMATION: a, t, c or g
; US-10-802-796-448

Query Match      66.4%; Score 14.6; DB 1; Length 295;
Best Local Similarity 81.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy      2  GAAAGTGATCGTCGCGCTGGT 22
Db      131 GAATGTGGTCGCTGGCGTCGT 111

RESULT 6
US-10-802-796-216/c
; Sequence 216, Application US/10802796
; Publication No. US20050250104A1
; GENERAL INFORMATION:
; APPLICANT: COLE, STEWART
; APPLICANT: BUCHRIESER-BROSCH, ROLAND
; APPLICANT: GORDON, STEPHEN
; APPLICANT: BILLAULT, ALAIN
; TITLE OF INVENTION: A METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST
; TITLE OF INVENTION: FROM THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED
; TITLE OF INVENTION: DNA LIBRARY. APPLICATION TO THE DETECTION OF
; TITLE OF INVENTION: MYCOBACTERIA.
; FILE REFERENCE: 05394.0011-00000
; CURRENT APPLICATION NUMBER: US/10/802,796
; CURRENT FILING DATE: 2004-03-18
; PRIOR APPLICATION NUMBER: US/09/673,476
; PRIOR FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: PCT/IB99/00740
; PRIOR FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 09/060,756
; PRIOR FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: PatentIn Ver. 2.2
; SEQ ID NO 216
; LENGTH: 325
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; US-10-802-796-216

Query Match      66.4%; Score 14.6; DB 1; Length 325;
Best Local Similarity 81.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy      2  GAAAGTGATCGTCGCGCTGGT 22
Db      96  GAATGTGGTCGCTGGCGTCGT 76

RESULT 7
US-10-802-796-496/c
; Sequence 496, Application US/10802796
; Publication No. US20050250104A1
; GENERAL INFORMATION:
; APPLICANT: COLE, STEWART
; APPLICANT: BUCHRIESER-BROSCH, ROLAND
; APPLICANT: GORDON, STEPHEN
; APPLICANT: BILLAULT, ALAIN
; TITLE OF INVENTION: A METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST
; TITLE OF INVENTION: FROM THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED
; TITLE OF INVENTION: DNA LIBRARY. APPLICATION TO THE DETECTION OF
; TITLE OF INVENTION: MYCOBACTERIA.
; FILE REFERENCE: 05394.0011-00000
; CURRENT APPLICATION NUMBER: US/10/802,796
; CURRENT FILING DATE: 2004-03-18
; PRIOR APPLICATION NUMBER: US/09/673,476
; PRIOR FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: PCT/IB99/00740
; PRIOR FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 09/060,756
; PRIOR FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: PatentIn Ver. 2.2
; SEQ ID NO 496
; LENGTH: 420
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
```

```

; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (21)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (23)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (106)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (200)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (272)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (355)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (362)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (413)
; OTHER INFORMATION: a, t, c or g
; OTHER INFORMATION: a, t, c or g
US-10-802-796-496

```

```

Query Match 66.4%; Score 14.6; DB 1; Length 420;
Best Local Similarity 81.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 2 GAAAGTGATCGCTGCCGTGGT 22
||| ||| ||| ||| ||| ||| |||
DB 128 GAATGTGCTGCTGGCGTGTCT 108

```

```

RESULT 8
US-11-082-389-413/c
; Sequence 413, Application US/11082389
; Publication No. US20050244935A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
; TITLE OF INVENTION: INVOLVED IN MEMBRANE SYNTHESIS AND MEMBRANE
; TITLE OF INVENTION: TRANSPORT
; FILE REFERENCE: BGI-131CPCN
; CURRENT APPLICATION NUMBER: US/11/082,389
; PRIOR FILING DATE: 2005-03-16
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/603024
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 60/141031
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: US 60/143262
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: US 60/151281
; PRIOR FILING DATE: 1999-07-01
; PRIOR APPLICATION NUMBER: DE 19930487.4
; PRIOR FILING DATE: 1999-07-01
; PRIOR APPLICATION NUMBER: DE 19930489.0
; PRIOR FILING DATE: 1999-07-01
; PRIOR APPLICATION NUMBER: DE 19931549.3
; PRIOR FILING DATE: 1999-07-08

```

```

; PRIOR APPLICATION NUMBER: DE 19931550.7
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19932134.5
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19941379.7
; PRIOR FILING DATE: 1999-08-31
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 446
; SEQ ID NO 413
; LENGTH: 428
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(405)
; OTHER INFORMATION: RXA02628
US-11-082-389-413

```

```

Query Match 66.4%; Score 14.6; DB 9; Length 428;
Best Local Similarity 81.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 2 GAAAGTGATCGCTGCCGTGGT 22
||| ||| ||| ||| ||| ||| |||
DB 237 GAAGTTGATCGCTGGGTGAT 217

```

```

RESULT 9
US-10-821-234-418
; Sequence 418, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 418
; LENGTH: 444
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-821-234-418

```

```

Query Match 66.4%; Score 14.6; DB 1; Length 444;
Best Local Similarity 81.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 1 AGAAAGTGATCGCTGCCGTGG 21
||| ||| ||| ||| ||| ||| |||
DB 395 AGAAAGTGCTGCTGGTGTGG 415

```

```

RESULT 10
US-10-802-796-75/c
; Sequence 75, Application US/10802796
; Publication No. US20050250104A1
; GENERAL INFORMATION:
; APPLICANT: COLE, STEWART
; APPLICANT: BUCHRIESER-BROSCH, ROLAND
; APPLICANT: GORDON, STEPHEN
; APPLICANT: BILLAULT, ALAIN
; TITLE OF INVENTION: A METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST
; TITLE OF INVENTION: FROM THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED
; TITLE OF INVENTION: DNA LIBRARY. APPLICATION TO THE DETECTION OF
; TITLE OF INVENTION: MYCOBACTERIA.
; FILE REFERENCE: 05394.0011-00000

```


; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 16138
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-16138

Query Match 64.5%; Score 14.2; DB 9; Length 19;
Best Local Similarity 63.2%; Pred. No. 1.4e+02;
Matches 12; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAAAGTATCGCTGCCGTG 20
|||||:|:|:|:|:
Db 1 GAAAGUGAUGCGCUAUG 19

RESULT 15
US-10-821-234-225
; Sequence 225, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 225
; LENGTH: 1018
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-821-234-225

Query Match 64.5%; Score 14.2; DB 1; Length 1018;
Best Local Similarity 84.2%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAGTATCGCTGCCGTGG 21
|||||:|:|:|:|:
Db 204 AAAGTGACAGTGCCTATGG 222

Search completed: November 27, 2005, 02:44:30
Job time : 7.76471 secs

RESULT 14
US-11-083-784-16138
; Sequence 16138, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:00:41 ; Search time 25.3623 Seconds
(without alignments)
7173.088 Million cell updates/sec

Title: US-09-555-529-18

Perfect score: 22

Sequence: 1 agaaagtgcgcgtgcgtggt 22

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA Main:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq*
- 2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq*
- 3: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq*
- 4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq*
- 5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq*
- 6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq*
- 7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq*
- 8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq*
- 9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq*
- 10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	22	100.0	588	4	US-09-925-065A-20834, A
C 2	17.4	79.1	381	3	US-09-732-627A-1180, Ap
C 3	17.4	79.1	1946	9	US-10-450-763-21563, A
C 4	17.4	79.1	3031	7	US-10-655-042-13, Appl
C 5	17.4	79.1	3307	7	US-10-655-042-15, Appl
C 6	17.4	79.1	3334	7	US-10-655-042-5, Appl
C 7	17.4	79.1	3388	7	US-10-655-042-7, Appl
C 8	17.4	79.1	3487	7	US-10-655-042-11, Appl
C 9	17.4	79.1	3532	7	US-10-655-042-3, Appl
C 10	17.4	79.1	3586	7	US-10-655-042-9, Appl
C 11	17.4	79.1	3661	7	US-10-633-577-1, Appl
C 12	17.4	79.1	3661	7	US-10-655-042-1, Appl
C 13	17.2	78.2	359	8	US-10-425-115-17122, A
C 14	17.2	78.2	412	3	US-09-864-761-21719, A
C 15	17.2	78.2	421	3	US-09-864-761-4990, A
C 16	17.2	78.2	475	7	US-10-296-115-58, Appl
C 17	17.2	78.2	555	6	US-10-029-385-5272, Ap
C 18	17.2	78.2	584	4	US-09-925-065A-475820, A
C 19	17.2	78.2	612	6	US-10-369-493-35681, A
C 20	17.2	78.2	754	3	US-09-910-943-130, Appl
C 21	17.2	78.2	1038	6	US-10-156-761-1058, Ap
C 22	17.2	78.2	1254	4	US-09-925-065A-56098, A
C 23	17.2	78.2	1254	4	US-09-925-065A-56099, A

C	24	17.2	78.2	1254	4	US-09-925-065A-56100	Sequence 56100, A
C	25	17.2	78.2	1348	7	US-10-437-963-36556	Sequence 36556, A
C	26	17.2	78.2	1867	7	US-10-302-172-749	Sequence 749, App
C	27	17.2	78.2	106664	5	US-10-175-523-97	Sequence 97, Appl
C	28	17.2	78.2	106664	10	US-11-099-266-97	Sequence 97, Appl
C	29	17.2	78.2	9025608	6	US-10-156-761-1	Sequence 1, Appl
C	30	16.8	76.4	156	7	US-10-437-963-16553	Sequence 16553, A
C	31	16.8	76.4	975	3	US-09-738-626-420	Sequence 420, App
C	32	16.8	76.4	1077	5	US-10-027-806-75	Sequence 75, Appl
C	33	16.8	76.4	1077	5	US-10-034-623-75	Sequence 75, Appl
C	34	16.8	76.4	1077	5	US-10-027-801-75	Sequence 75, Appl
C	35	16.8	76.4	1077	6	US-10-029-120-75	Sequence 75, Appl
C	36	16.8	76.4	1886	10	US-11-097-143-10247	Sequence 10247, A
C	37	16.8	76.4	2107	7	US-10-425-114-13350	Sequence 13350, A
C	38	16.8	76.4	2115	7	US-10-424-599-11032	Sequence 11032, A
C	39	16.8	76.4	4153	10	US-11-097-143-10246	Sequence 10246, A
C	40	16.8	76.4	32998	5	US-10-027-806-1	Sequence 1, Appl
C	41	16.8	76.4	32998	5	US-10-034-623-1	Sequence 1, Appl
C	42	16.8	76.4	32998	5	US-10-027-801-1	Sequence 1, Appl
C	43	16.8	76.4	32998	6	US-10-029-120-1	Sequence 1, Appl
C	44	16.8	76.4	238417	9	US-10-461-862-98	Sequence 98, Appl
C	45	16.8	76.4	3309400	3	US-09-738-626-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-925-065A-20834/c
; Sequence 20834, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925.065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20834
; LENGTH: 588
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-20834

Query Match 100.0%; Score 22; DB 4; Length 588;
Best Local Similarity 100.0%; Pred. No. 0.44; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0;

QY 1 AGAAAGTGCCTGCGTGTGT 22
|||||
Db 67 AGAAAGTGCCTGCGTGTGT 46
|||||

RESULT 2
US-09-732-627A-1180/c
; Sequence 1180, Application US/09732627A
; Publication No. US2004012338A1
; GENERAL INFORMATION:
; APPLICANT: Fincher, Karen L.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants

```
; FILE REFERENCE: 38-21(51770)B
; CURRENT APPLICATION NUMBER: US/09/732,627A
; CURRENT FILING DATE: 2000-12-08
; NUMBER OF SEQ ID NOS: 4930
; SEQ ID NO 1180
; LENGTH: 381
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3493-057-P1-M1-D2
US-09-732-627A-1180

Query Match          79.1%; Score 17.4; DB 3; Length 381;
Best Local Similarity 94.7%; Pred. No. 98;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGAAGTGCCTGCGCGGT 19
    |||||
Db 37 AGAAGTGAACGTCGCGT 19

RESULT 3
US-10-450-763-21563/c
; Sequence 21563, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 21563
; LENGTH: 1946
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIMILAR
; LOCATION: (944)..(1165)
; OTHER INFORMATION: 94% homologous to Escherichia coli tref, accession number
; OTHER INFORMATION: U00039, Smith-Waterman Score=369.
US-10-450-763-21563

Query Match          79.1%; Score 17.4; DB 9; Length 1946;
Best Local Similarity 94.7%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 AAGTATCGCTGCGCGGT 22
    |||||
Db 219 AAGTATCGCTGCGCGGT 201

RESULT 4
US-10-655-042-13
; Sequence 13, Application US/10655042
; Publication No. US20040126842A1
; GENERAL INFORMATION:
; APPLICANT: CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
; TITLE OF INVENTION: MUTANT E. COLI STRAINS, AND THEIR USE FOR PRODUCING RECOMBINANT
; FILE REFERENCE: ECO
; CURRENT APPLICATION NUMBER: US/10/655,042
; CURRENT FILING DATE: 2003-09-05
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13

Query Match          79.1%; Score 17.4; DB 7; Length 3307;
Best Local Similarity 94.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 AAGTATCGCTGCGCGGT 22
    |||||
Db 2833 AAGTATCGCTGCGCGGT 2851

RESULT 5
US-10-655-042-15
; Sequence 15, Application US/10655042
; Publication No. US20040126842A1
; GENERAL INFORMATION:
; APPLICANT: CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
; TITLE OF INVENTION: MUTANT E. COLI STRAINS, AND THEIR USE FOR PRODUCING RECOMBINANT
; FILE REFERENCE: ECO
; CURRENT APPLICATION NUMBER: US/10/655,042
; CURRENT FILING DATE: 2003-09-05
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 3307
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: E. coli RNase E deletion mutant
; NAME/KEY: CDS
; LOCATION: (441)..(3269)
; OTHER INFORMATION:
US-10-655-042-15

Query Match          79.1%; Score 17.4; DB 7; Length 3307;
Best Local Similarity 94.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 AAGTATCGCTGCGCGGT 22
    |||||
Db 2833 AAGTATCGCTGCGCGGT 2851

RESULT 6
US-10-655-042-5
; Sequence 5, Application US/10655042
; Publication No. US20040126842A1
; GENERAL INFORMATION:
; APPLICANT: CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
; TITLE OF INVENTION: MUTANT E. COLI STRAINS, AND THEIR USE FOR PRODUCING RECOMBINANT
; FILE REFERENCE: ECO
; CURRENT APPLICATION NUMBER: US/10/655,042
; CURRENT FILING DATE: 2003-09-05
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 3334
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: E. coli RNase E deletion mutant
```

```
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (441)..(3296)
; OTHER INFORMATION:
US-10-655-042-5

Query Match      79.1%; Score 17.4; DB 7; Length 3334;
Best Local Similarity 94.7%; Pred.No.1.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 AAGTGATCGCTGCCGTGGT 22
        |||||
DB       2860 AAGTGATCGCTGCCGCGGT 2878

RESULT 7
US-10-655-042-7
; Sequence 7, Application US/10655042
; Publication No. US20040126842A1
; GENERAL INFORMATION:
; APPLICANT: CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
; TITLE OF INVENTION: MUTANT E. COLI STRAINS, AND THEIR USE FOR PRODUCING RECOMBINANT
; TITLE OF INVENTION: POLYPEPTIDES
; FILE REFERENCE: ECO
; CURRENT APPLICATION NUMBER: US/10/655,042
; CURRENT FILING DATE: 2003-09-05
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 3388
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: E. coli RNase E deletion mutant
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (441)..(3350)
; OTHER INFORMATION:
US-10-655-042-7

Query Match      79.1%; Score 17.4; DB 7; Length 3388;
Best Local Similarity 94.7%; Pred.No.1.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 AAGTGATCGCTGCCGTGGT 22
        |||||
DB       2914 AAGTGATCGCTGCCGCGGT 2932

RESULT 8
US-10-655-042-11
; Sequence 11, Application US/10655042
; Publication No. US20040126842A1
; GENERAL INFORMATION:
; APPLICANT: CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
; TITLE OF INVENTION: MUTANT E. COLI STRAINS, AND THEIR USE FOR PRODUCING RECOMBINANT
; TITLE OF INVENTION: POLYPEPTIDES
; FILE REFERENCE: ECO
; CURRENT APPLICATION NUMBER: US/10/655,042
; CURRENT FILING DATE: 2003-09-05
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 3487
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: E. coli RNase E deletion mutant
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (441)..(3449)
; OTHER INFORMATION:
US-10-655-042-11
```

QY 4 AAGTGATCGCTGCCGTGGT 22
|||||
Db 3112 AAGTGATCGCTGCCGGGT 3130

RESULT 11
US-10-633-577-1
; Sequence 1, Application US/10633577
; Publication No. US20040029231A1
; GENERAL INFORMATION:
; APPLICANT: DREYFUS, MARC
; APPLICANT: LOPEZ, PASCAL
; TITLE OF INVENTION: MUTANT E. COLI STRAINS, AND THEIR USE FOR PRODUCING
; TITLE OF INVENTION: RECOMBINANT POLYPEPTIDES
; FILE REFERENCE: US98APNREC
; CURRENT APPLICATION NUMBER: US/10/633,577
; PRIOR FILING DATE: 2003-08-05
; PRIOR APPLICATION NUMBER: PCT/FR99/01879
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: FR 98/10197
; PRIOR FILING DATE: 1998-08-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 3661
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (441)..(3623)
US-10-633-577-1

Query Match 79.1%; Score 17.4; DB 7; Length 3661;
Best Local Similarity 94.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4 AAGTGATCGCTGCCGTGGT 22
|||||
Db 3187 AAGTGATCGCTGCCGGGT 3205

RESULT 12
US-10-655-042-1
; Sequence 1, Application US/10655042
; Publication No. US20040126842A1
; GENERAL INFORMATION:
; APPLICANT: CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
; TITLE OF INVENTION: MUTANT E. COLI STRAINS, AND THEIR USE FOR PRODUCING RECOMBINANT
; TITLE OF INVENTION: POLYPEPTIDES
; FILE REFERENCE: ECO
; CURRENT APPLICATION NUMBER: US/10/655,042
; CURRENT FILING DATE: 2003-09-05
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 3661
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (441)..(3623)
; OTHER INFORMATION:
US-10-655-042-1

Query Match 79.1%; Score 17.4; DB 7; Length 3661;
Best Local Similarity 94.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4 AAGTGATCGCTGCCGTGGT 22
|||||
Db 3187 AAGTGATCGCTGCCGGGT 3205

RESULT 13
US-10-425-115-17122
; Sequence 17122, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 17122
; LENGTH: 359
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_115611C.1
US-10-425-115-17122

Query Match 78.2%; Score 17.2; DB 8; Length 359;
Best Local Similarity 86.4%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AGAAAGTGATCGCTGCCGTGGT 22
|||||
Db 239 AGAAAGTGATCGCTGCCGTGAT 260

RESULT 14
US-09-864-761-21719/c
; Sequence 21719, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661

```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 21719
; LENGTH: 412
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL031665.18
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 25
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 27
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 19
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 30
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 26
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 45
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 38
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 33
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 44
; OTHER INFORMATION: NT HIT: D50657.1, EVALUATION 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P53506, EVALUATION 2.00e-52
; OTHER INFORMATION: EST_HUMAN HIT: BE616766.1, EVALUATION 5.00e-93
US-09-864-761-21719
```

```
Query Match 78.2%; Score 17.2; DB 3; Length 412;
Best Local Similarity 86.4%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGAAGATGATCGCTGCGCTGGT 22
    ||||| ||||| ||||| |||||
DB 168 AGAAGAGATCACTGCGCTGGT 147
```

```
RESULT 15
US-09-864-761-4990/c
; Sequence 4990, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wengheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aomic-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
```

```
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 4990
; LENGTH: 421
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL031665.18
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 25
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 27
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 19
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 30
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 26
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 45
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 38
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 33
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 44
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 44
US-09-864-761-4990
```

```
Query Match 78.2%; Score 17.2; DB 3; Length 421;
Best Local Similarity 86.4%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 AGAAGATGATCGCTGCGCTGGT 22
    ||||| ||||| ||||| |||||
DB 291 AGAAGAGATCACTGCGCTGGT 270
```

Search completed: November 27, 2005, 02:22:34
Job time : 32.3623 secs

This Page Blank (usp10)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OW nucleic - nucleic search, using SW model

Run on: November 26, 2005, 13:00:02 ; Search time 4.556 Seconds
(without alignments)
8583.479 Million cell updates/sec

Title: US-09-555-529-18

Sequence: 1 agaaagtcgctgcctgcgtggt 22

Scoring table: IDENTITY NUC
Gapex 10.0, Gapex 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1-COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5-COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A-COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B-COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H-COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCUS-COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP-COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE-COMB.seq:*
9: /cgn2_6/ptodata/1/ina/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	100.0	23645	US-09-949-016-13916	Sequence 13916, A
2	17.4	79.1	3661	US-09-762-481B-1	Sequence 1, Appl1
3	17.2	78.2	1867	US-09-799-451-749	Sequence 749, App
4	16.8	76.4	968	US-09-605-703B-2171	Sequence 2171, Ap
5	16.8	76.4	1077	US-09-408-020-75	Sequence 75, Appl
6	16.8	76.4	32998	US-09-408-020-1	Sequence 1, Appl1
7	16.2	73.6	501	US-09-252-991A-13073	Sequence 13073, A
8	16.2	73.6	601	US-09-949-016-63710	Sequence 63710, A
9	16.2	73.6	605	US-08-630-915A-5	Sequence 5, Appl1
10	16.2	73.6	605	US-09-879-957-5	Sequence 5, Appl1
11	16.2	73.6	2586	US-09-107-532A-2231	Sequence 2331, Ap
12	16.2	73.6	2994	US-09-949-016-1858	Sequence 1858, Ap
13	16.2	73.6	15449	US-09-949-016-13600	Sequence 13600, A
14	16.2	73.6	42325	US-08-311-731A-111	Sequence 131, App
15	16.2	73.6	340380	US-09-949-016-14179	Sequence 14179, A
16	15.8	71.8	479	US-09-270-767-4961	Sequence 4961, Ap
17	15.8	71.8	479	US-09-270-767-20243	Sequence 20243, A
18	15.8	71.8	715	US-08-998-416-590	Sequence 590, App
19	15.8	71.8	4494	US-09-620-312D-861	Sequence 861, App
20	15.8	71.8	6263	US-08-781-802-3	Sequence 3, Appl1
21	15.8	71.8	26711	US-08-694-078-3	Sequence 3, Appl1
22	15.8	71.8	42931	US-09-949-016-14214	Sequence 14214, A
23	15.8	71.8	42931	US-08-311-731A-129	Sequence 129, App
24	15.8	71.8	4403765	US-09-103-840A-2	Sequence 2, Appl1

25	15.8	71.8	441529	3	US-09-103-840A-1	Sequence 1, Appl1
26	15.6	70.9	435	3	US-09-902-540-4149	Sequence 4149, Ap
27	15.6	70.9	437	3	US-09-270-767-11731	Sequence 11731, A
28	15.6	70.9	485	3	US-09-513-999C-13085	Sequence 13085, A
29	15.6	70.9	541	3	US-09-270-767-8408	Sequence 8408, Ap
30	15.6	70.9	541	3	US-09-270-767-23690	Sequence 23690, A
31	15.6	70.9	588	3	US-08-858-207A-250	Sequence 250, App
32	15.6	70.9	601	3	US-09-949-016-122813	Sequence 122813, A
33	15.6	70.9	601	3	US-09-949-016-130765	Sequence 130765, A
34	15.6	70.9	601	3	US-09-949-002-3379	Sequence 3379, Ap
35	15.6	70.9	601	3	US-09-949-002-9273	Sequence 9273, Ap
36	15.6	70.9	671	3	US-09-533-559-6307	Sequence 6307, Ap
37	15.6	70.9	732	3	US-09-252-991A-14582	Sequence 14582, A
38	15.6	70.9	1002	3	US-09-543-681A-447	Sequence 447, App
39	15.6	70.9	1356	3	US-09-270-767-10484	Sequence 10484, A
40	15.6	70.9	1401	3	US-09-489-039A-483	Sequence 483, App
41	15.6	70.9	1488	3	US-08-962-859A-1	Sequence 1, Appl1
42	15.6	70.9	1488	3	US-09-583-110-1905	Sequence 1905, Ap
43	15.6	70.9	1488	3	US-09-107-433-785	Sequence 785, App
44	15.6	70.9	1828	3	US-09-664-958-11	Sequence 11, Appl
45	15.6	70.9	1866	4	US-09-605-703B-603	Sequence 603, App

ALIGNMENTS

```

RESULT 1
US-09-949-016-13916/C
; Sequence 13916, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 13916
; LENGTH: 23645
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-13916

Query Match      100.0%; Score 22; DB 3; Length 23645;
Best Local Similarity 100.0%; Pred. No. 0.44;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGAAAGTATCGCTGCGCTGCT 22
      |||||
DB      1789 AGAAAGTATCGCTGCGCTGCT 1768

RESULT 2
US-09-762-481B-1
; Sequence 1, Application US/09762481B
; Patent No. 6632639
; GENERAL INFORMATION:
; APPLICANT: DREYERUS, MARC
; APPLICANT: LOPEZ, PASCAL
; TITLE OF INVENTION: MUTANT E. COLI STRAINS, AND THEIR USE FOR PRODUCING
; RECOMBINANT POLYPEPTIDES
; FILE REFERENCE: US898APCNRCC
; CURRENT APPLICATION NUMBER: US/09/762,481B
; CURRENT FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: PCT/FR99/01879

```

```
;; PRIOR FILING DATE: 1999-07-29
;; PRIOR APPLICATION NUMBER: FR 98/10197
;; PRIOR FILING DATE: 1998-08-07
;; NUMBER OF SEQ ID NOS: 6
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 1
;; LENGTH: 3661
;; TYPE: DNA
;; ORGANISM: Escherichia coli
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (441)..(3623)
US-09-762-481B-1
```

```
Query Match          79.1%; Score 17.4; DB 3; Length 3661;
Best Local Similarity 94.7%; Pred. No. 64;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY      4 AAGTATCGCTGCCGTGT 22
          |||||
Db      3187 AAGTATCGCTGCCGTGT 3205
```

```
RESULT 3
US-09-799-451-749
;; Sequence 749, Application US/09799451
;; Patent No. 6783969
;; GENERAL INFORMATION:
;; APPLICANT: Tang, Y. Tom
;; APPLICANT: Zhou, Ping
;; APPLICANT: Goodrich, Ryle
;; APPLICANT: Asundi, Vinod
;; APPLICANT: Ren, Feiyan
;; APPLICANT: Zhang, Jie
;; APPLICANT: Xue, Aidong J.
;; APPLICANT: Zhao, Qing A.
;; APPLICANT: Wang, Jian-Rui
;; APPLICANT: Ma, Yundong
;; APPLICANT: Yamazaki, Victoria
;; APPLICANT: Chen, Rui-hong
;; APPLICANT: Wang, Zhiwei
;; APPLICANT: Wang, Dunfui
;; APPLICANT: Yang, Yonghong
;; APPLICANT: Wehrman, Tom
;; APPLICANT: Ghosh, Reena
;; APPLICANT: Drmanac, Radjce T.
;; TITLE OF INVENTION: No. 6783969e1 Nucleic Acids and
;; FILE REFERENCE: 803
;; CURRENT APPLICATION NUMBER: US/09/799,451
;; CURRENT FILING DATE: 2001-03-05
;; NUMBER OF SEQ ID NOS: 948
;; SOFTWARE: pc_fl_genes Version 2.0
;; SEQ ID NO 749
;; LENGTH: 1867
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1103)..(1675)
US-09-799-451-749
```

```
Query Match          78.2%; Score 17.2; DB 3; Length 1867;
Best Local Similarity 86.4%; Pred. No. 73;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
OY      1 AGAAGTATCGCTGCCGTGT 22
          |||||
Db      1628 AGAAGTATCGCTGCCGTGT 1649
```

```
RESULT 4
US-09-605-703B-2171
```

```
;; Sequence 2171, Application US/09605703B
;; Patent No. 6962989
;; GENERAL INFORMATION:
;; APPLICANT: Pompejus, Markus
;; APPLICANT: Kroger, Burkhard
;; APPLICANT: Schroder, Hartwig
;; APPLICANT: Zelder, Oskar
;; APPLICANT: Haberhauer, Gregor
;; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING NOVEL
;; FILE REFERENCE: BGI-129CP
;; CURRENT APPLICATION NUMBER: US/09/605,703B
;; CURRENT FILING DATE: 2000-06-27
;; PRIOR APPLICATION NUMBER: 60/142,764
;; PRIOR FILING DATE: 1999-07-08
;; PRIOR APPLICATION NUMBER: 60/152,318
;; PRIOR FILING DATE: 1999-09-03
;; NUMBER OF SEQ ID NOS: 2934
;; SEQ ID NO 2171
;; LENGTH: 968
;; TYPE: DNA
;; ORGANISM: Corynebacterium glutamicum
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)..(945)
;; OTHER INFORMATION: RXA01313
US-09-605-703B-2171
```

```
Query Match          76.4%; Score 16.8; DB 4; Length 968;
Best Local Similarity 90.0%; Pred. No. 1,1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
OY      1 AGAAGTATCGCTGCCGTG 20
          |||||
Db      665 AGAAGCATCGCTGCCGTG 684
```

```
RESULT 5
US-09-408-020-75/c
;; Sequence 75, Application US/09408020
;; Patent No. 6632937
;; GENERAL INFORMATION:
;; APPLICANT: Swanson, Ronald V.
;; APPLICANT: Feldman, Robert A.
;; APPLICANT: Schleper, Christa
;; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
;; FILE REFERENCE: DORP.002A
;; CURRENT APPLICATION NUMBER: US/09/408,020
;; CURRENT FILING DATE: 1999-09-29
;; PRIOR APPLICATION NUMBER: 60/102,294
;; PRIOR FILING DATE: 1998-09-29
;; NUMBER OF SEQ ID NOS: 123
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 75
;; LENGTH: 1077
;; TYPE: DNA
;; ORGANISM: Cenarchaeum symbiosum
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)....(1077)
US-09-408-020-75
```

```
Query Match          76.4%; Score 16.8; DB 3; Length 1077;
Best Local Similarity 90.0%; Pred. No. 1,1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
OY      3 AAAGTATCGCTGCCGTGT 22
          |||||
Db      828 AAAGTCTCTGCTGCCGTGT 809
```

```
RESULT 6
US-09-408-020-1
```

Sequence 1, Application US/09408020
Patent No. 6632937
GENERAL INFORMATION:
APPLICANT: Swanson, Ronald V.
APPLICANT: Feldman, Robert A.
TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
FILE REFERENCE: DCOIP.002A
CURRENT APPLICATION NUMBER: US/09/408,020
CURRENT FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: 60/102,294
PRIOR FILING DATE: 1998-09-29
NUMBER OF SEQ ID NOS: 123
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 32998
TYPE: DNA
ORGANISM: Cenarchaeum symbiosum
FEATURE:
NAME/KEY: CDS
LOCATION: (7604)...(8908)
FEATURE:
NAME/KEY: CDS
LOCATION: (8961)...(9767)
FEATURE:
NAME/KEY: CDS
LOCATION: (10545)...(10922)
FEATURE:
NAME/KEY: CDS
LOCATION: (13944)...(14612)
FEATURE:
NAME/KEY: CDS
LOCATION: (18638)...(20149)
FEATURE:
NAME/KEY: CDS
LOCATION: (20554)...(20955)
FEATURE:
NAME/KEY: CDS
LOCATION: (20956)...(21834)
FEATURE:
NAME/KEY: CDS
LOCATION: (25151)...(26377)
FEATURE:
NAME/KEY: CDS
LOCATION: (27535)...(28002)
FEATURE:
NAME/KEY: CDS
LOCATION: (28065)...(29483)
US-09-408-020-1

Query Match 76.4%; Score 16.8; DB 3; Length 32998;
Best Local Similarity 90.0%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AAGTATCGCTGCGCGTGT 22
DB 26627 AAGTCTGCTGCTGCGTGT 26646

RESULT 7
US-09-252-991A-13073
Sequence 13073, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 13073
LENGTH: 501
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13073

Query Match 73.6%; Score 16.2; DB 3; Length 501;
Best Local Similarity 85.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGAAGTATCGCTGCGCGTGT 21
DB 251 AGAAGTATCGCTGCGTGT 271

RESULT 8
US-09-949-016-63710/c
Sequence 63710, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CLO01307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 63710
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-63710

Query Match 73.6%; Score 16.2; DB 3; Length 601;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGAAGTATCGCTGCGCGTGT 21
DB 486 AGAGGCGATGCTGCGCGTGT 466

RESULT 9
US-08-630-915A-5/c
Sequence 5, Application US/08630915A
Patent No. 6309820
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
APPLICANT: HOFMAN, No. 6309820h
APPLICANT: KAY, Brian K.
APPLICANT: FOWLES, Dana M.
APPLICANT: MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSES: Pennie & Edwards LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Mastrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 605 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-630-915A-5

Query Match 73.6%; Score 16.2; DB 3; Length 605;
Best Local Similarity 85.7%; Pred. No. 2e+02; 3; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GAAAGTATCGCTGCCGTGCT 22
Db 431 GCAAGTGTTCGCTGCCGGGT 411

RESULT 10
US-09-879-957-5/c
Sequence 5, Application US/09879957
Patent No. 6709821
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
HOFFMAN, No. 6709821h
KAY, Brian K.
FOWLES, Dana M.
MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
USING SAME
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/879,957
FILING DATE: 13-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/630,915
FILING DATE: 03-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Mastrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 605 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-879-957-5

Query Match 73.6%; Score 16.2; DB 3; Length 605;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GAAAGTATCGCTGCCGTGCT 22
Db 431 GCAAGTGTTCGCTGCCGGGT 411

RESULT 11
US-09-107-532A-2331/c
Sequence 2331, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneka
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 2331:
SEQUENCE CHARACTERISTICS:
LENGTH: 2586 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...2586
SEQUENCE DESCRIPTION: SEQ ID NO: 2331:
US-09-107-532A-2331

Query Match 73.6%; Score 16.2; DB 3; Length 2586;
Best Local Similarity 85.7%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAAAGTATCGCTGCCGTGT 22
DB 1122 GAAAGTATCGCTGCCGTGT 1102

RESULT 12
US-09-949-016-1858
Sequence 1858, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1858
LENGTH: 2994
TYPE: DNA
ORGANISM: Human
US-09-949-016-1858

Query Match 73.6%; Score 16.2; DB 3; Length 2994;
Best Local Similarity 85.7%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGAAGTATCGCTGCCGTGT 21
DB 2317 AGAAGTATCGCTGCCGTGT 2337

RESULT 13
US-09-949-016-13600
Sequence 13600, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13600
LENGTH: 15449
TYPE: DNA
ORGANISM: Human
US-09-949-016-13600

Query Match 73.6%; Score 16.2; DB 3; Length 15449;
Best Local Similarity 85.7%; Pred. No. 3.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGAAGTATCGCTGCCGTGT 21
DB 12771 AGAAGTATCGCTGCCGTGT 12791

RESULT 14
US-08-311-731A-131/C
Sequence 131, Application US/08311731A
Patent No. 6583266
GENERAL INFORMATION:
APPLICANT: SMITH, DOUGLAS
APPLICANT: MAO, JEN-I
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
DIAGNOSTICS AND THERAPEUTICS
TITLE OF INVENTION: 411
NUMBER OF SEQUENCES: 411
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P. C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311,731A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: C0044/7125
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/720-3500
TELEFAX: 617/720-2441
INFORMATION FOR SEQ ID NO: 131:
SEQUENCE CHARACTERISTICS:
LENGTH: 42325 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: MYCOBACTERIUM LEPRAE
US-08-311-731A-131

Query Match 73.6%; Score 16.2; DB 3; Length 42325;
Best Local Similarity 85.7%; Pred. No. 3.5e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGAAGTATCGCTGCCGTGT 21
DB 15586 AGGAGTATCGCTGCCGTGT 15566

RESULT 15
US-09-949-016-14179
Sequence 14179, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755

```
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14179
; LENGTH: 340380
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1) .. (340380)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14179
```

```
Query Match      73.6%; Score 16.2; DB 3; Length 340380;
Best Local Similarity 85.7%; Pred. No. 4.5e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      2  GAAAGTATCGCTGCCCTGGT  22
          |||||  |||||  |||||
Db      219584  GAAAGTGACAGCTGCCCTGGT  219604
```

```
Search completed: November 27, 2005, 01:10:47
Job time : 10.556 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 12:58:53 : Search time 122,977 Seconds
(without alignments)
8370.011 Million cell updates/sec

Title: US-09-555-529-18
Perfect score: 22
Sequence: 1 agaaatgacgcgtgcgtgggt 22

Scoring table: IDENTITY_NUC
Gapop 10,0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*

1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	100.0	529	5	BX504040 DKFZp686H
2	22	100.0	581	3	BP350798 BP350798
3	22	100.0	718	8	CX866441 HESCA 10
4	22	100.0	882	3	BQ221694 AGENCOURT
5	22	100.0	890	6	CD109094 AGENCOURT
6	22	100.0	922	6	CD389584 AGENCOURT
7	22	100.0	965	1	AL522388 AL522388
8	22	100.0	1339	4	CR618602 full-length
9	21	95.5	552	7	CN256426 CN256426
10	20.4	92.7	951	7	C0725903 C0725903
11	20.4	92.7	988	7	C0726806 C0726806
12	20.4	92.7	1011	7	C0725532 C0725532
13	19.4	88.2	411	5	BY450707 BY450707
14	19	86.4	677	8	DR001039 DR001039
15	18	81.8	842	3	BT862461 BT862461
16	17.8	80.9	106	9	BH818205 BH818205
17	17.8	80.9	310	9	BH817324 BH817324
18	17.8	80.9	365	8	CV987860 CV987860
19	17.8	80.9	370	5	BW888824 BW888824
20	17.8	80.9	453	7	CN211166 Rt fcd 04
21	17.8	80.9	508	9	BH825901 BH825901
22	17.8	80.9	529	5	BU697159 BU697159

23	17.8	80.9	530	10	CZ633455	CZ633455 OM_Ba017
24	17.8	80.9	554	9	BH828298	BH828298 BACPP27-D
25	17.8	80.9	591	10	CZ611198	CZ611198 OM_Ba013
26	17.8	80.9	633	8	CV993795	CV993795 IPCGF+3.1
27	17.8	80.9	649	6	CB939232	CB939232 IPCGJ+13
28	17.8	80.9	650	8	CV996304	CV996304 IPCGF+4.1
29	17.8	80.9	660	6	CB937334	CB937334 IPCGJ+13
30	17.8	80.9	668	7	CF947745	CF947745 UI-D-GC0-
31	17.8	80.9	694	8	CV994840	CV994840 IPCGF+3.1
32	17.8	80.9	713	5	BU221080	BU221080 603755374
33	17.8	80.9	713	9	BH835596	BH835596 BACPP8-H1
34	17.8	80.9	754	10	CM673179	CM673179 OB_Ba006
35	17.8	80.9	763	3	BJ050799	BJ050799 BJ050799
36	17.8	80.9	776	10	CZ675529	CZ675529 OM_Ba023
37	17.8	80.9	789	5	BU450710	BU450710 603217766
38	17.8	80.9	896	10	CNS02CW4	AL191677 Tetraodon
39	17.8	80.9	925	7	CK410458	CK410458 AUF_IpHdk
40	17.4	79.1	115	6	CF557778	CF557778 1115031C0
41	17.4	79.1	116	6	CF561674	CF561674 1115067D1
42	17.4	79.1	163	7	CV406745	CV406745 RCO-TN008
43	17.4	79.1	395	1	AU050627	AU050627 AU050627
44	17.4	79.1	432	5	BY461925	BY461925 BY461925
45	17.4	79.1	543	10	CL978385	CL978385 OsIFCC031

ALIGNMENTS

RESULT 1
BX504040
LOCUS
DEFINITION BX504040 529 bp mRNA linear EST 04-SEP-2003
DKFZp686H04126 r1 686 (synonym: hlcc3) Homo sapiens CDNA clone
DKFZp686H04126 5', mRNA sequence.
ACCESSION BX504040
VERSION BX504040.1 GI:32028491
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 529)
AUTHORS Poustka A., Albert R., Moosmayer P., Schupp I., Wellenreuther R.,
Mewes H.W., Weil B., Amid C., Osanger A., Fobo G., Han M. and
Wiemann S.
TITLE EST (Poustka A., Albert R., Moosmayer P., Schupp I.,
Wellenreuther R., et al.)
JOURNAL Unpublished (2003)
COMMENT Contact: MIPS
MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by DKFZ (German Cancer Research Center,
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
No sl sequence available.
This clone (DKFZp686H04126) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES

Location/Qualifiers
1..529
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686H04126"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: hlcc3)"
/note="vector: pTriplex2; Site_1: sfIIA; Site_2: SfiIB;
cDNA-collection"

ORIGIN

```

Query Match      100.0%; Score 22; DB 5; Length 529;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAGTGCCTGCGCGTGGT 22
    |||||
Db 36 AGAAGTGCCTGCGCGTGGT 57

RESULT 2
LOCUS BP350798
DEFINITION BP350798 Sugano cDNA library, squamous cell TE13 Homo sapiens cDNA
clone T3R01993, mRNA sequence.
ACCESSION BP350798
VERSION BP350798.1 GI:52280784
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Suzuki,Y., Yamashita,R., Shiota,M., Sakakibara,Y., Chiba,J.,
AUTHORS Mizushima-Sugano,J., Nakai,K. and Sugano,S.
TITLE Sequence comparison of human and mouse genes reveals a homologous
JOURNAL block structure in the promoter regions
PUBMED Genome Res. 14 (9), 1711-1718 (2004)
COMMENT 15342556
Contact: Yutaka Suzuki
Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Email: yusuzuki@ims.u-tokyo.ac.jp.

FEATURES
source
1..581
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="T3R01993"
/cell_type="squamous cell"
/cell_line="TE13"
/clone_lib="Sugano cDNA library, squamous cell TE13"
/notes="well-differentiated squamous cell carcinoma"

ORIGIN
Query Match      100.0%; Score 22; DB 3; Length 581;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAGTGCCTGCGCGTGGT 22
    |||||
Db 26 AGAAGTGCCTGCGCGTGGT 47

RESULT 3
LOCUS CX866441
DEFINITION HESCA_10_H01.G1_A037 NIH_MGC_262 Homo sapiens cDNA clone
IMAGE:7472092 5', mRNA sequence.
ACCESSION CX866441
VERSION CX866441.1 GI:58548219
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

Other_ESTs: HESCA_10_H01.B1_A037
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-remail.nih.gov
Tissue Procurement: BresaGen, Inc.
cDNA Library Preparation: Express Genomics, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Laboratory for Genomics and Bioinformatics,
University of Georgia
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM15776 row: 0 column: 02
Seq primer: JENREV (CAGGAACAGCATGACC)
High quality sequence stop: 718.

FEATURES
source
1..718
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:7472092"
/sex="male"
/tissue_type="embryonic stem"
/cell_type="human embryonic stem cells"
/cell_line="BG01"
/lab_host="DH10B-T1 phage-resistant E. coli"
/clone_lib="NIH_MGC_262"
/notes="Vector: pExpress-1; Site 1: NotI; Site 2: EcoRV;
RNA obtained from human embryonic stem cells isolated from
the inner cell mass of blastocyst stage embryos and
differentiated to an early neural progenitor cell type.
Cell line id and NIH Registry designation is BG01.
Positive for Nestin and Musashi expression. Passage number
18. cDNA primed using oligo-dT primer:
5'-GCATGATCTAGATCGGAGCGCGGCC(T)25-3' and cloned into
the EcoRV/NotI sites of pExpress-1. This primary library
is non-normalized (normalized primary library is
NIH_MGC_259). It was constructed by Express Genomics
(Frederick, MD). Sequence ends have been trimmed to
exclude vector and regions below phred quality 16. Note:
this is a Mammalian Gene Collection library."

ORIGIN
Query Match      100.0%; Score 22; DB 8; Length 718;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAGTGCCTGCGCGTGGT 22
    |||||
Db 2 AGAAGTGCCTGCGCGTGGT 23

RESULT 4
LOCUS BQ221694
DEFINITION AGENCOURT_7549812 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6057750
5', mRNA sequence.
ACCESSION BQ221694
VERSION BQ221694.1 GI:20403094
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Mammalia; Eutheria; Chordata; Vertebrata; Euteleostomi;
AUTHORS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Hominidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

```

Tissue Procurement: DCTD/DTP/Gazdar
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LNA1321 row: 1 column: 07
 High quality sequence stop: 581.
 Location/Qualifiers
 1. :882
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6057750"
 /tissue_type="large cell carcinoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC 68"
 /note="Organ: lung; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.8 kb. Library constructed by Life Technologies."

ORIGIN

Query Match 100.0%; Score 22; DB 3; Length 882;
 Best Local Similarity 100.0%; Pred. No. 7;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGAAGTGCCTGCGCGTGGT 22
 |||||
 Db 5 AGAAGTGCCTGCGCGTGGT 26

RESULT 5

CD109094
 LOCUS
 DEFINITION AGENCOURT 14019646 NIH_MGC_179 Homo sapiens cDNA clone
 IMAGE:30366888 5', mRNA sequence.
 CD109094
 VERSION
 KEYWORDS EST.
 SOURCE
 ORGANISM Homo sapiens (human)

REFERENCE

1 (bases 1 to 890)
 NIH-MGC <http://mgi.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. Michael Brownstein
 cDNA Library Preparation: Invitrogen Corp
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: NDA437 row: d column: 01
 High quality sequence stop: 653.
 Location/Qualifiers
 1. :890
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30366888"
 /tissue_type="Pituitary"
 /lab_host="DH10B-Ton A (T1 and T5 phage resistances)"
 /clone_lib="NIH_MGC 179"
 /note="Organ: brain; Vector: pCMV-SPORT6.1; Site 1: EcoRV (destroyed); Site 2: NotI; Library is oligo-dt primed and directionally cloned (EcoRV site is destroyed upon

FEATURES

source

cloning). Average insert size 1.1 kb. Library was constructed by (Invitrogen). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 890;
 Best Local Similarity 100.0%; Pred. No. 7;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGAAGTGCCTGCGCGTGGT 22
 |||||
 Db 22 AGAAGTGCCTGCGCGTGGT 43

RESULT 6

CD389584
 LOCUS
 DEFINITION AGENCOURT 14304876 NIH_MGC_173 Homo sapiens cDNA 5', mRNA sequence.
 CD389584
 ACCESSION
 VERSION
 KEYWORDS EST.
 SOURCE
 ORGANISM Homo sapiens (human)

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homnidae; Homo.

REFERENCE
 1 (bases 1 to 922)
 NIH-MGC <http://mgi.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. Jamie Thompson, University of WI
 cDNA Library Preparation: Gina Zastrow-Hayes
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: NDKM67 row: n column: 11
 High quality sequence start: 28
 High quality sequence stop: 601.
 Location/Qualifiers
 1. :922
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /tissue_type="embryonic trophoblasts, made from WA01 stem cells"

FEATURES

source
 1. :922
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /tissue_type="embryonic trophoblasts, made from WA01 stem cells"
 /lab_host="DH10B Tona"
 /clone_lib="NIH_MGC 173"
 /note="Vector: pDONR201; Site 1: attP2; Site 2: attP1;
 LIBR PRIMING - oligo dt; METHOD - full-length enriched;
 LIBR PROVIDER - Bradfield"

ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 922;
 Best Local Similarity 100.0%; Pred. No. 7;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGAAGTGCCTGCGCGTGGT 22
 |||||
 Db 38 AGAAGTGCCTGCGCGTGGT 59

RESULT 7

AL522388
 LOCUS
 DEFINITION AL522388 Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED Homo sapiens cDNA clone CS0DB008YGl8 5-PRIME, mRNA sequence.

```

ACCESSION      AL522388
VERSION        AL522388.3  GI:45697719
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
               Homnidae; Homo.
REFERENCE      1 (bases 1 to 965)
AUTHORS        Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE          Full-length cDNA libraries and normalization
JOURNAL        Unpublished (2001)
COMMENT        On Feb 13, 2001 this sequence version replaced gi:31040660.
               Contact: Genoscope
               Genoscope - Centre National de Sequencage
               2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
               Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
               1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
               end enriched, double-strand cDNA was digested with Not I and cloned
               into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
               was normalized. Library was constructed by Life Technologies, a
               division of Invitrogen. This sequence belongs to sequence cluster
               5543.r
               For more information about this cluster, see
               http://www.genoscope.cns.fr/cdna?s=CS0DB008BD09QP1&c=5543.r.
               Location/Qualifiers
               1..965
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /clone="CS0DB008YGI8"
               /tissue_type="NEUROBLASTOMA COT 10-NORMALIZED"
               /clone_lib="Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED"
               /notes="1st strand cDNA was primed with a NotI-oligo(dT)
               primer. Five prime end enriched, double-strand cDNA was
               digested with Not I and EcoR V sites of the pCMVSPORT 6 vector.
               Library was normalized."
ORIGIN
Query Match      100.0%; Score 22; DB 1; Length 965;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAAGTGCCTGCGCGTGGT 22
    |||||
Db 41 AGAAAGTGCCTGCGCGTGGT 62

RESULT 8
LOCUS           CR618602
DEFINITION      full-length cDNA clone CS0DB008YGI8 of Neuroblastoma Cot
ACCESSION       CR618602.1  GI:50499409
VERSION         CR618602.1
KEYWORDS        HTC; CNSLT_CDNA.
SOURCE          Homo sapiens
ORGANISM        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
               Homnidae; Homo.
REFERENCE      1 (bases 1 to 1339)
AUTHORS        Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE          Full-length cDNA libraries and normalization
JOURNAL        Unpublished
REMARK         Contact : Feng Liang Email : fliang@lifetech.com URL :
               http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
               Faraday Avenue
               2 (bases 1 to 1339)
               Genoscope.
               Direct Submission
               Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
               BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

ACCESSION      AL522388
VERSION        AL522388.3  GI:45697719
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
               Homnidae; Homo.
REFERENCE      1 (bases 1 to 965)
AUTHORS        Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE          Full-length cDNA libraries and normalization
JOURNAL        Unpublished (2001)
COMMENT        On Feb 13, 2001 this sequence version replaced gi:31040660.
               Contact: Genoscope
               Genoscope - Centre National de Sequencage
               2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
               Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
               1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
               end enriched, double-strand cDNA was digested with Not I and cloned
               into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
               was normalized. Library was constructed by Life Technologies, a
               division of Invitrogen. This sequence belongs to sequence cluster
               5543.r
               For more information about this cluster, see
               http://www.genoscope.cns.fr/cdna?s=CS0DB008BD09QP1&c=5543.r.
               Location/Qualifiers
               1..965
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /clone="CS0DB008YGI8"
               /tissue_type="NEUROBLASTOMA COT 10-NORMALIZED"
               /clone_lib="Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED"
               /notes="1st strand cDNA was primed with a NotI-oligo(dT)
               primer. Five prime end enriched, double-strand cDNA was
               digested with Not I and EcoR V sites of the pCMVSPORT 6 vector.
               Library was normalized."
ORIGIN
Query Match      100.0%; Score 22; DB 1; Length 965;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAAGTGCCTGCGCGTGGT 22
    |||||
Db 41 AGAAAGTGCCTGCGCGTGGT 62

RESULT 8
LOCUS           CR618602
DEFINITION      full-length cDNA clone CS0DB008YGI8 of Neuroblastoma Cot
ACCESSION       CR618602.1  GI:50499409
VERSION         CR618602.1
KEYWORDS        HTC; CNSLT_CDNA.
SOURCE          Homo sapiens
ORGANISM        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
               Homnidae; Homo.
REFERENCE      1 (bases 1 to 1339)
AUTHORS        Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE          Full-length cDNA libraries and normalization
JOURNAL        Unpublished
REMARK         Contact : Feng Liang Email : fliang@lifetech.com URL :
               http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
               Faraday Avenue
               2 (bases 1 to 1339)
               Genoscope.
               Direct Submission
               Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
               BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
Location/Qualifiers
1..1339
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DB008YGI8"
/tissue_type="Neuroblastoma Cot 10-normalized"
/plasmid="pCMVSPORT_6"
FEATURES
source
ORIGIN
Query Match      100.0%; Score 22; DB 4; Length 1339;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAAGTGCCTGCGCGTGGT 22
    |||||
Db 41 AGAAAGTGCCTGCGCGTGGT 62

RESULT 9
LOCUS           CN256426
DEFINITION      17000600085458 GRN_PREHEP Homo sapiens cDNA 5', mRNA sequence.
ACCESSION       CN256426
VERSION         CN256426.1  GI:47272840
KEYWORDS        EST.
SOURCE          Homo sapiens (human)
ORGANISM        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
               Homnidae; Homo.
REFERENCE      1 (bases 1 to 552)
AUTHORS        Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G.J.,
               Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M.S., Mandalam, R.,
               Lebkowski, J. and Stanton, L.W.
TITLE          Transcriptome characterization elucidates signaling networks that
               control human ES cell growth and differentiation
JOURNAL        Nat. Biotechnol. 22 (6), 707-716 (2004)
PUBMED         15146197
COMMENT        Contact: Brandenberger R
               Regenerative Medicine
               Genon Corporation
               230 Constitution Drive, Menlo Park, CA 94025, USA
               Tel: 650 473 8658
               Fax: 650 473 7760
               Email: rbrandenberger@genon.com
               Insert Length: 552 Std Error: 0.00.
               Location/Qualifiers
               1..552
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /tissue_type="embryonic stem cells, DMSO-treated H9 cell
               line"
               /clone_lib="GRN PREHEP"
               /note="oligo dT primed, full-length enriched cDNA library
               from DMSO-treated HES cell line H9 (p22) maintained in
               feeder-free conditions"
FEATURES
source
ORIGIN
Query Match      95.5%; Score 21; DB 7; Length 552;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GAAAGTGCCTGCGCGTGGT 22
    |||||
Db 9 GAAAGTGCCTGCGCGTGGT 29

```

```

RESULT 10
C0725903
LOCUS
DEFINITION
ILLUMIGEN MCQ 40291 Katze MMPB2 Macaca mulatta cDNA clone
IBUW:27173 5' similar to Bases 5 to 950 highly similar to human
KIN (Hs.123647), mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Macaca mulatta (rhesus monkey)

REFERENCE
AUTHORS
Magnes, C.L., Fellin, P.C., Thomas, M.J., Korth, M.J., Agt, M.B.,
Prohl, S.C., Fitzgibbon, M., Scherer, C.A., Miner, D.G., Katze, M.G. and
Iadonato, S.P.

TITLE
Analysis of the Macaca mulatta transcriptome and the sequence
divergence between Macaca and human

JOURNAL
PUBMED
Genome Biol. 6 (7), R60 (2005)

COMMENT
Contact: C. Magnes
ILLUMIGEN Biosciences Inc.
2203 Airport Way S, Suite 450, Seattle, WA 98134, USA
Tel: 2063780400
Fax: 2063780408
Email: cmagnes@illumigen.com
Sequenced on 2004.05.28 723 Q20 bases. Library Preparation: Prof.
Michael Katze Lab at University of Washington DNA Sequencing:
ILLUMIGEN Biosciences Inc. For further information, see
http://www.macaque.org

PCR Primers
FORWARD: CCCTCACTAAAGGGAACAAAA
BACKWARD: CACTATAGGCGAATTGGGTA
Insert Length: 951 Std Error: 0.00
Plate: CL000321 row: C column: 12
Seq primer: CCCTCACTAAAGGGAACAAAA
POLYA-No.

FEATURES
source
1..951
/organism="Macaca mulatta"
/mol_type="mRNA"
/strain="Indian"
/db_xref="taxon:9544"
/clone="IBUW:27173"
/sex="male"
/tissue_type="blood"
/cell_type="PBMC"
/dev_stage="adult"
/lab_host="Electromax DH10B"
/clone_lib="Katze MMPB2"
/notes="Vector: pDONR 222; Site 1: Berg I; Site 2: Berg I;
Created from Cloneminer cDNA Library Construction kit
(catalog #18249-029)"

ORIGIN
Query Match 92.7%; Score 20.4; DB 7; Length 951;
Best Local Similarity 95.5%; Pred. No. 44;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGAAGTGTATCGCTGCCGTGGT 22
|||||
Db 27 AGAAGTGTATCGCTGCCGTGGT 48
|||||

RESULT 11
C0726806
LOCUS
DEFINITION
ILLUMIGEN MCQ 48930 Katze WMDD Macaca mulatta cDNA clone
IBUW:27517 5' similar to Bases 5 to 967 highly similar to human
KIN (Hs.123647), mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Macaca mulatta (rhesus monkey)

REFERENCE
AUTHORS
Magnes, C.L., Fellin, P.C., Thomas, M.J., Korth, M.J., Agt, M.B.,
Prohl, S.C., Fitzgibbon, M., Scherer, C.A., Miner, D.G., Katze, M.G. and
Iadonato, S.P.

TITLE
Analysis of the Macaca mulatta transcriptome and the sequence
divergence between Macaca and human

JOURNAL
PUBMED
Genome Biol. 6 (7), R60 (2005)

COMMENT
Contact: C. Magnes
ILLUMIGEN Biosciences Inc.
2203 Airport Way S, Suite 450, Seattle, WA 98134, USA
Tel: 2063780400
Fax: 2063780408
Email: cmagnes@illumigen.com
Sequenced on 2004.05.28 723 Q20 bases. Library Preparation: Prof.
Michael Katze Lab at University of Washington DNA Sequencing:
ILLUMIGEN Biosciences Inc. For further information, see
http://www.macaque.org

PCR Primers
FORWARD: CCCTCACTAAAGGGAACAAAA
BACKWARD: CACTATAGGCGAATTGGGTA
Insert Length: 951 Std Error: 0.00
Plate: CL000321 row: C column: 12
Seq primer: CCCTCACTAAAGGGAACAAAA
POLYA-No.

FEATURES
source
1..951
/organism="Macaca mulatta"
/mol_type="mRNA"
/strain="Indian"
/db_xref="taxon:9544"
/clone="IBUW:27173"
/sex="male"
/tissue_type="blood"
/cell_type="PBMC"
/dev_stage="adult"
/lab_host="Electromax DH10B"
/clone_lib="Katze WMDD"
/notes="Vector: pDONR 222; Site 1: Berg I; Site 2: Berg I;
Created from Cloneminer cDNA Library Construction kit
(catalog #18249-029)"

ORIGIN
Query Match 92.7%; Score 20.4; DB 7; Length 951;
Best Local Similarity 95.5%; Pred. No. 44;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGAAGTGTATCGCTGCCGTGGT 22
|||||
Db 27 AGAAGTGTATCGCTGCCGTGGT 48
|||||

RESULT 12
C0725532
LOCUS
DEFINITION
ILLUMIGEN MCQ 36075 Katze MMBR Macaca mulatta cDNA clone
IBUW:28218 5' similar to Bases 165 to 958 highly similar to human
KIN (Hs.123647), mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Macaca mulatta (rhesus monkey)

REFERENCE
AUTHORS
Magnes, C.L., Fellin, P.C., Thomas, M.J., Korth, M.J., Agt, M.B.,
Prohl, S.C., Fitzgibbon, M., Scherer, C.A., Miner, D.G., Katze, M.G. and
Iadonato, S.P.

TITLE
Analysis of the Macaca mulatta transcriptome and the sequence
divergence between Macaca and human

JOURNAL
PUBMED
Genome Biol. 6 (7), R60 (2005)

COMMENT
Contact: C. Magnes
ILLUMIGEN Biosciences Inc.
2203 Airport Way S, Suite 450, Seattle, WA 98134, USA
Tel: 2063780400
Fax: 2063780408
Email: cmagnes@illumigen.com
Sequenced on 2004.07.02 706 Q20 bases. Library Preparation: Prof.
Michael Katze Lab at University of Washington DNA Sequencing:
ILLUMIGEN Biosciences Inc. For further information, see
http://www.macaque.org

PCR Primers
FORWARD: CCCTCACTAAAGGGAACAAAA
BACKWARD: CACTATAGGCGAATTGGGTA
Insert Length: 988 Std Error: 0.00
Plate: CL000405 row: C column: 07
Seq primer: CCCTCACTAAAGGGAACAAAA
POLYA=yes.

FEATURES
source
1..988
/organism="Macaca mulatta"
/mol_type="mRNA"
/strain="Indian"
/db_xref="taxon:9544"
/clone="IBUW:27517"
/sex="male"
/dev_stage="adult"
/lab_host="Electromax DH10B"
/clone_lib="Katze WMDD"
/notes="Organ: duodenum; Vector: pDONR 222; Site 1: Berg I;
Site 2: Berg I; Created from Cloneminer cDNA Library
Construction kit (catalog #18249-029)"

ORIGIN
Query Match 92.7%; Score 20.4; DB 7; Length 988;
Best Local Similarity 95.5%; Pred. No. 45;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGAAGTGTATCGCTGCCGTGGT 22
|||||
Db 6 AGAAGTGTATCGCTGCCGTGGT 27
|||||

RESULT 12
C0725532
LOCUS
DEFINITION
ILLUMIGEN MCQ 36075 Katze MMBR Macaca mulatta cDNA clone
IBUW:28218 5' similar to Bases 165 to 958 highly similar to human
KIN (Hs.123647), mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Macaca mulatta (rhesus monkey)

REFERENCE
AUTHORS
Magnes, C.L., Fellin, P.C., Thomas, M.J., Korth, M.J., Agt, M.B.,
Prohl, S.C., Fitzgibbon, M., Scherer, C.A., Miner, D.G., Katze, M.G. and
Iadonato, S.P.

TITLE
Analysis of the Macaca mulatta transcriptome and the sequence
divergence between Macaca and human

JOURNAL
PUBMED
Genome Biol. 6 (7), R60 (2005)

COMMENT
Contact: C. Magnes
ILLUMIGEN Biosciences Inc.
2203 Airport Way S, Suite 450, Seattle, WA 98134, USA
Tel: 2063780400
Fax: 2063780408
Email: cmagnes@illumigen.com
Sequenced on 2004.07.02 706 Q20 bases. Library Preparation: Prof.
Michael Katze Lab at University of Washington DNA Sequencing:
ILLUMIGEN Biosciences Inc. For further information, see
http://www.macaque.org

PCR Primers
FORWARD: CCCTCACTAAAGGGAACAAAA
BACKWARD: CACTATAGGCGAATTGGGTA
Insert Length: 988 Std Error: 0.00
Plate: CL000405 row: C column: 07
Seq primer: CCCTCACTAAAGGGAACAAAA
POLYA=yes.

FEATURES
source
1..988
/organism="Macaca mulatta"
/mol_type="mRNA"
/strain="Indian"
/db_xref="taxon:9544"
/clone="IBUW:27517"
/sex="male"
/dev_stage="adult"
/lab_host="Electromax DH10B"
/clone_lib="Katze WMDD"
/notes="Organ: duodenum; Vector: pDONR 222; Site 1: Berg I;
Site 2: Berg I; Created from Cloneminer cDNA Library
Construction kit (catalog #18249-029)"

ORIGIN
Query Match 92.7%; Score 20.4; DB 7; Length 988;
Best Local Similarity 95.5%; Pred. No. 45;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGAAGTGTATCGCTGCCGTGGT 22
|||||
Db 6 AGAAGTGTATCGCTGCCGTGGT 27
|||||

```

REFERENCE AUTHORS	1 (bases 1 to 1011) Magness,C.L., Fellin,P.C., Thomas,M.J., Korth,M.J., Agy,M.B., Prolli,S.C., Fitzgibbon,M., Scherer,C.A., Miner,D.G., Katze,M.G. and Iadonato,S.P.
TITLE	Analysis of the Macaca mulatta transcriptome and the sequence divergence between Macaca and human
JOURNAL PUBMED	Genome Biol. 6 (7), R60 (2005) 15998449
COMMENT	Contact: C. Magness Illumigen Biosciences Inc. 2203 Airport Way S, Suite 450, Seattle, WA 98134, USA Tel: 2063780400 Fax: 2063780408 Email: cmagness@illumigen.com Sequenced on 2004.05.11. 622 Q20 bases. Library Preparation: Prof. Michael Katze Lab at University of Washington DNA Sequencing: Illumigen Biosciences Inc. For further information, see http://www.macaque.org PCR Primers FORWARD: CCTCTACTAAGGGAACAAAA BACKWARD: CACTATAGGCGAATTTGGGTA Insert Length: 1011 Std Error: 0.00 Plate: CL000250 row: E column: 09 Seq primer: CCCTCACTAAGGGAACAAAA POLYA=Yes.
FEATURES source	Location/Qualifiers 1..1011 /organism="Macaca mulatta" /mol_type="mRNA" /strain="Indian" /db_xref="taxon:9544" /clone="IBIUM:28218" /sex="female" /dev_stage="adult" /lab_host="E. coli SOLR" /clone_lib="Katze WMBR" /notes="Organ: brain; Vector: Uni-ZAP XR; Site: 1: Ecor I; Site: 2: Xho I; created from Stragene ZAP-CDNA Synthesis kit (Catalog #200400) and ZAP-CDNA Gigapack III Gold Cloning kit (Catalog #200450)"
ORIGIN	Query Match 92.7%; Score 20.4; DB 7; Length 1011; Best Local Similarity 95.5%; Pred. No. 45; Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy	1 AGAAAGTGATCGTCGCGTGGT 22
Db	25 AGAAAGTGATCGTCGCGTGGT 46
RESULT 13 BY450707 LOCUS	BY450707 411 bp mRNA linear EST 13-DEC-2002
DEFINITION	musculus cDNA clone K630047B05 3', mRNA sequence.
ACCESSION VERSION KEYWORDS SOURCE	BY450707 BY450707.1 GI:26744401 EST. Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 411) Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
REFERENCE AUTHORS	Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Guscinich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Sempile,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Vardaro,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Harai,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002) 12466851 Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/ Aizawa,K., Akimura,I., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001) Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site (http://genome.gsc.riken.go.jp) for further details. Location/Qualifiers 1..411 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="K630047B05" /tissue_type="whole body" /dev_stage="14 days embryo" /clone_lib="RIKEN full-length enriched, 14 days embryo whole body"
ORIGIN	Query Match 88.2%; Score 19.4; DB 5; Length 411; Best Local Similarity 95.2%; Pred. No. 1.2e+02; Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy	1 AGAAAGTGATCGTCGCGTGG 21

```

Db      31 AGAAGTGATAGCTGCGCTGG 51
|||||
RESULT 14
LOCUS   DR001039
DEFINITION
TC112658 Human fetal brain, large insert, pCMV expression library
Homo sapiens cDNA clone TC112658 5' similar to Homo sapiens KIN,
antigenic determinant of recA protein homolog (mouse) (KIN), mRNA
sequence.
ACCESSION
DR001039
VERSION  DR001039.1  GI:66260912
KEYWORDS
SOURCE   Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE
1 (bases 1 to 677)
Birkett,C., Cho,J., Gau,Y., Hamer,R., Kelly,S., Kovacs,K., Liu,L.,
Liu,X., Porter,J., Sachs,A., Shu,Y., Sun,Z., Wong,J., Wu,M.,
Zhang,X., Jay,G. and He,W.
High-throughput cloning of full-length human cDNAs directly from
cDNA libraries optimized for large and rare transcripts
Unpublished (2005)
CONTACT: Kovacs, KF
High Throughput cDNA Cloning
Origene Technologies, Inc. ( www.origene.com )
6 Taft Court, Suite 100, Rockville, MD 20850, USA
Tel: 301 340 3188
Fax: 301 340 8606
Email: cDNA@origene.com
This EST submission is part of an on-going human full-length
cloning project at Origene Technologies, Inc.
Please contact Origene for access.
Origene Technologies, Inc.
6 Taft Ct. Suite 100
Rockville, MD 20850
Tel: (301) 340-3188
http://www.origene.com
Seq primer: pCMV6 5prime forward vector primer, Origene
Technologies Inc.
FEATURES             source
Location/Qualifiers
1..677
/organism="Homo sapiens"
/mol_type="mRNA"
/db|xref="taxon:9606"
/clone="TC112658"
/tissue_type="Fetal Brain"
/clone_lib="Human fetal brain, large insert, pCMV
expression library"
/notes="Organ: Fetal Brain; Vector: pCMV6-XL4; Site 1:
EcoRI; Site 2: XhoI/SalI compatible end ligatio; Oligo-dT
primed reverse transcription optimized for large and GC
rich mRNA transcripts, cDNA size selection, optimized
ligation for large inserts into mammalian expression
vector, random clones selected for end sequence
verification of full-length genes"
ORIGIN
Query Match      86.4%; Score 19; DB 8; Length 677;
Best Local Similarity 100.0%; Pred.No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4 AAGTGATCGCTGCGCTGGT 22
|||||
Db      26 AAGTGATCGCTGCGCTGGT 44
|||||

RESULT 15
LOCUS   BI862461
DEFINITION
603391378F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5406499 5',
mRNA sequence.
ACCESSION
BI862461
VERSION  BI862461.1  GI:16003208
KEYWORDS
SOURCE   Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE
1 (bases 1 to 842)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
CONTACT: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue procurement: DCTD/DTF
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM12036 row: 1 column: 20
High quality sequence stop: 834.
FEATURES             source
Location/Qualifiers
1..842
/organism="Homo sapiens"
/mol_type="mRNA"
/db|xref="taxon:9606"
/clone="IMAGE:5406499"
/tissue_type="mammary adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_87"
/notes="Organ: Breast; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.383 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
ORIGIN
Query Match      81.8%; Score 18; DB 3; Length 842;
Best Local Similarity 100.0%; Pred.No. 6.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 AGAAGTGATCGCTGCGCG 18
|||||
Db      23 AGAAGTGATCGCTGCGCG 40
|||||

Search completed: November 27, 2005, 00:57:53
Job time : 125.977 secs

```


This Page Blank (usp1c)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 12:48:57 ; Search time 14.1112 Seconds
(without alignments)
10390.557 Million cell updates/sec

Title: US-09-555-529-18
Perfect score: 22
Sequence: 1 agaaagtgcgtgcgtggt 22

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21.*

1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001as.*
5: Geneseq2001bs.*
6: Geneseq2002as.*
7: Geneseq2002bs.*
8: Geneseq2003as.*
9: Geneseq2003bs.*
10: Geneseq2003cs.*
11: Geneseq2003ds.*
12: Geneseq2004as.*
13: Geneseq2004bs.*
14: Geneseq2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	22	2	AAx85566 PCR prime
2	22	100.0	1002	2	AAx85551 CDNA of a
3	22	100.0	1296	2	AAx85549 CDNA of a
C 4	17.4	79.1	381	12	ADP92169 Cotton ex
C 5	17.4	79.1	1946	5	AAx85759 DNA encod
6	17.4	79.1	3661	3	AAx291119 Mutated E
7	17.4	79.1	3661	3	AAx291118 Escherich
C 8	17.2	78.2	412	4	AAx24336 Probe #14
C 9	17.2	78.2	412	4	ABA69474 Human fo
C 10	17.2	78.2	412	4	AAx49606 Probe #18
C 11	17.2	78.2	412	4	ABA51446 Human bre
C 12	17.2	78.2	412	4	ABA36399 Probe #14
C 13	17.2	78.2	412	4	AAx43550 Human bon
C 14	17.2	78.2	412	4	AAx17729 Human bra
C 15	17.2	78.2	412	4	ABA43195 Human liv
C 16	17.2	78.2	412	5	AAx09878 Probe #98
C 17	17.2	78.2	412	6	ABA17689 Human gen
C 18	17.2	78.2	421	4	AAx15156 Probe #50
C 19	17.2	78.2	421	4	ABA56909 Human fo

C 20	17.2	78.2	421	4	AAx136484	Aa136484 Probe #51
C 21	17.2	78.2	421	4	ABA46344	ABA46344 Human bre
C 22	17.2	78.2	421	4	ABA26524	ABA26524 Probe #49
C 23	17.2	78.2	421	4	AAK30540	Aak30540 Human bon
C 24	17.2	78.2	421	4	AAK05006	Aak05006 Human bra
C 25	17.2	78.2	421	4	ABS30203	ABs30203 Human liv
C 26	17.2	78.2	421	5	AAx04899	AAx04899 Probe #48
C 27	17.2	78.2	421	6	ABS05185	ABs05185 Human gen
C 28	17.2	78.2	421	4	AAH9223	AAH9223 Human pro
C 29	17.2	78.2	475	14	AEA20832	Aea20832 Novel hum
C 30	17.2	78.2	555	12	ACH72077	Ach72077 Human gen
C 31	17.2	78.2	612	13	ADS60007	AdS60007 Bacterial
C 32	17.2	78.2	754	6	ABs76882	ABs76882 Frog embr
C 33	17.2	78.2	1867	6	ABZ11867	Abz11867 Human pol
C 34	17.2	78.2	1867	12	ADM44385	AdM44385 Novel hum
C 35	17.2	78.2	106664	13	ADV35021	Adv35021 Murine cd
C 36	16.8	76.4	457	4	AAK56789	Aak56789 Human imm
C 37	16.8	76.4	975	5	AAH65385	Aah65385 C glutami
C 38	16.8	76.4	975	8	ACA02093	AcA02093 C. glutam
C 39	16.8	76.4	1077	3	AAA55224	Aaa55224 C. symbio
C 40	16.8	76.4	1081	4	AAK68843	Aak68843 Human imm
C 41	16.8	76.4	1081	4	AAK68842	Aak68842 Human imm
C 42	16.8	76.4	1886	4	ABL08671	AbL08671 Drosophil
C 43	16.8	76.4	2107	13	ADX30530	Adx30530 Plant ful
C 44	16.8	76.4	4153	4	ABL08670	AbL08670 Drosophil
C 45	16.8	76.4	32998	3	AAA55186	Aaa55186 Cenarchae

ALIGNMENTS

RESULT 1

AAx85566

ID AAx85566 standard; cDNA; 22 BP.

XX AAx85566;

DT 07-SEP-1999 (first entry)

DE PCR primer and probe used to detect human kin17 nucleic acids.

KW Human; kin17 protein; cell proliferation; fertility; probe;

KW hyperproliferative disease; protein interaction; curved DNA;

KW HIV replication; HIV integration; repair enzyme; PCR primer; ss.

XX Synthetic.

OS Homo sapiens.

PN FR2772046-A1.

XX 11-JUN-1999.

PF 09-DEC-1997; 97FR-00015536.

XX 09-DEC-1997; 97FR-00015536.

XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;

XX WPI, 1999-359999/31.

XX New DNA coding for human kin17 protein - useful for controlling cell

XX proliferation or fertility.

XX Claim 7; Page 5; 69pp; French.

XX AAx8553-69 represent PCR primers and probes used to detect nucleic acids

XX encoding human kin17 protein. The mammalian kin17 protein is useful for

XX preparing a medicament for controlling cell proliferation or for

XX controlling fertility. The medicaments can also be used to treat

XX hyperproliferative diseases. Fragments between amino acids 55 and 235

XX (preferably between amino acids 129 and 228) of a mammalian kin17 protein

CC are useful for regulating the interaction between proteins and curved
 CC DNA. The fragment can be used to block replication of HIV or its
 CC integration into the human genome or to target repair enzymes to curved
 CC DNA sites. Expression vectors for kin17 can be used for controlling cell
 CC proliferation

SQ Sequence 22 BP; 5 A; 4 C; 8 G; 5 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 2; Length 22;
 Best Local Similarity 100.0%; Pred. NO. 0.49;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGAAGTGCCTGCGCGTGT 22
 |||||
 Db 1 AGAAGTGCCTGCGCGTGT 22

RESULT 2

AAx85551
 ID AAX85551 standard; cDNA; 1002 BP.

XX AC AAX85551;

DT 07-SEP-1999 (first entry)

XX cDNA of a gene coding for a truncated human kin17 protein.

XX kin17 protein; cell proliferation; fertility; hyperproliferative disease;
 KW protein interaction; curved DNA; HIV replication; HIV integration;
 KW repair enzyme; ss.

XX Homo sapiens.

OS FR2772046-A1.

XX 11-JUN-1999.

XX 09-DEC-1997; 97FR-00015536.

XX 09-DEC-1997; 97FR-00015536.

XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;

DR WPI; 1999-359999/31.

XX New DNA coding for human kin17 protein - useful for controlling cell
 PT proliferation or fertility.

XX Claim 5; Page 32; 69pp; French.

XX The present sequence encodes a truncated human kin17 protein with amino
 CC acids 129-228 deleted. The mammalian kin17 protein is useful for
 CC preparing a medicament for controlling cell proliferation or for
 CC controlling fertility. The medicaments can also be used to treat
 CC hyperproliferative diseases. Fragments between amino acids 55 and 235
 CC (preferably between amino acids 129 and 228) of a mammalian kin17 protein
 CC are useful for regulating the interaction between proteins and curved
 CC DNA. The fragment can be used to block replication of HIV or its
 CC integration into the human genome or to target repair enzymes to curved
 CC DNA sites. Expression vectors for kin17 can be used for controlling cell
 CC proliferation

XX Sequence 1002 BP; 343 A; 189 C; 230 G; 240 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 2; Length 1002;
 Best Local Similarity 100.0%; Pred. NO. 0.81;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGAAGTGCCTGCGCGTGT 22
 |||||
 Db 32 AGAAGTGCCTGCGCGTGT 53

RESULT 3
 AAX85549
 ID AAX85549 standard; cDNA; 1296 BP.

XX AC AAX85549;

XX 07-SEP-1999 (first entry)

XX cDNA of a gene coding for the human kin17 protein.

XX Human; kin17 protein; cell proliferation; fertility;
 KW hyperproliferative disease; protein interaction; curved DNA;
 KW HIV replication; HIV integration; repair enzyme; ss.

XX Homo sapiens.

XX FR2772046-A1.

XX 11-JUN-1999.

XX 09-DEC-1997; 97FR-00015536.

XX 09-DEC-1997; 97FR-00015536.

XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;

DR WPI; 1999-359999/31.

XX New DNA coding for human kin17 protein - useful for controlling cell
 PT proliferation or fertility.

XX Claim 1; Page 30-31; 69pp; French.

XX The present sequence encodes a human kin17 protein. The mammalian kin17
 CC protein is useful for preparing a medicament for controlling cell
 CC proliferation or for controlling fertility. The medicaments can also be
 CC used to treat hyperproliferative diseases. Fragments between amino acids
 CC 55 and 235 (preferably between amino acids 129 and 228) of a mammalian
 CC kin17 protein are useful for regulating the interaction between proteins
 CC and curved DNA. The fragment can be used to block replication of HIV or
 CC its integration into the human genome or to target repair enzymes to
 CC curved DNA sites. Expression vectors for kin17 can be used for
 CC controlling cell proliferation

SQ Sequence 1296 BP; 453 A; 240 C; 310 G; 293 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 2; Length 1296;
 Best Local Similarity 100.0%; Pred. NO. 0.84;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGAAGTGCCTGCGCGTGT 22
 |||||
 Db 32 AGAAGTGCCTGCGCGTGT 53

RESULT 4

ADP92169/c
 ID ADP92169 standard; cDNA; 381 BP.

XX AC ADP92169;

XX 09-SEP-2004 (first entry)

XX Cotton expressed sequence tag, EST, #1180.

XX Cotton; ss; EST; expressed sequence tag; plant; plant protection;
 KW plant improvement; marker-assisted breeding.

XX Gossypium hirsutum; variety Nucleon33B.

```

XX PN US2004123338-A1.
XX PD 24-JUN-2004.
XX PF
XX PP
XX PR 08-DEC-2000; 2000US-00732627.
XX PR 10-DEC-1999; 99US-0170255P.
XX PA (FINC/) FINCHER K L.
XX PI Fincher KL;
XX PI WPI; 2004-479807/45.
XX DR
XX PT New substantially purified nucleic acid molecule that encodes a cotton
XX PT protein or its fragment, useful as molecular tool for the targeting and
XX PT isolation of novel genes for plant protection and improvement.
XX PS Claim 1; SEQ ID NO 1180; 30pp; English.
XX CC The invention relates to a substantially purified nucleic acid molecule
XX CC that encodes a cotton protein or its fragment comprising an EST
XX CC (expressed sequence tag) appearing as ADP9090-ADP95919. Also included
XX CC are a substantially purified cotton protein or its fragment encoded by a
XX CC nucleic acid molecule above and a transformed plant (having a nucleic
XX CC acid molecule which comprises: an exogenous promoter region which
XX CC functions in a plant cell to cause the production of a mRNA molecule; a
XX CC structural nucleic acid molecule comprising one of the ESTs or their
XX CC complements; a 3' non-translated sequence that functions in the plant
XX CC cell to cause termination of transcription and addition of polyadenylated
XX CC ribonucleotides to a 3' end of the mRNA molecule). The ESTs are useful as
XX CC molecular tool for the targeting and isolation of novel genes for plant
XX CC protection and improvement. The ESTs are useful for developing new
XX CC strategies for understanding critical plant developmental and metabolic
XX CC pathways, for isolating genes and promoters, for identifying and mapping
XX CC the genes involved in developmental and metabolic pathways, and for
XX CC determining gene function. The cotton nucleic acid molecules are useful
XX CC as molecular tags to isolate genetic regions, isolate genes, map genes,
XX CC and determine gene function. The nucleic acid molecules are useful for
XX CC determining if genes are members of a particular gene family and for use
XX CC in marker-assisted breeding programs. The present sequence is one of the
XX CC 4930 cotton ESTs of the invention. NOTE: The sequences are not displayed
XX CC in the specification but are available in electronic format from the
XX CC USPTO at seqdata.uspto.gov/sequence.html?docID=20040123338.
XX SQ Sequence 381 BP; 99 A; 85 C; 114 G; 83 T; 0 U; 0 Other;

Query Match 79.1%; Score 17.4; DB 12; Length 381;
Best Local Similarity 94.7%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGAAGTGTATCGCTGCCGT 19
DB 37 AGAAGTGTATCGCTGCCGT 19

RESULT 5
ID AAS85759/c
XX ID AAS85759 standard; cDNA; 1946 BP.
XX AC AAS85759;
XX DT 13-FEB-2002 (first entry)
XX DE DNA encoding novel human diagnostic protein #21563.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX OS Homo sapiens.
XX FH WO200175067-A2.
XX FT

Query Match 79.1%; Score 17.4; DB 5; Length 1946;
Best Local Similarity 94.7%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 AAGTGATCGCTGCCGTGCT 22
DB 219 AAGTGATCGCTGCCGTGCT 201

RESULT 6
AAZ91119
ID AAZ91119 standard; DNA; 3661 BP.
XX AC AAZ91119;
XX DT 06-JUN-2000 (first entry)
XX DE Mutated Escherichia coli RNase E gene.
XX KW Ribonuclease E; RNase E; mutation; mRNA degradation; wild type;
XX KW protein truncation; recombinant gene expression; ds.
XX OS Escherichia coli.
XX OS Synthetic.
XX FH Key
XX FH Location/Qualifiers
XX FT 441..2198

```

```

XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PP 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSB-) HYSB INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX PI WPI; 2001-639362/73.
XX DR P-PSDB; ABG21572.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 1; SEQ ID NO 21563; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX CC sequences. (I) is useful as hybridisation probes, polymerase chain
XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX CC and in recombinant production of (II). The polynucleotides are also used
XX CC in diagnostics as expressed sequence tags for identifying expressed
XX CC genes. (I) is useful in gene therapy techniques to restore normal
XX CC activity of (II) or to treat disease states involving (II). (I) is
XX CC useful for generating antibodies against it, detecting or quantitating a
XX CC polypeptide in tissue, as molecular weight markers and as a food
XX CC supplement. (II) and its binding partners are useful in medical imaging
XX CC of sites expressing (II). (I) and (II) are useful for treating disorders
XX CC involving aberrant protein expression or biological activity. The
XX CC polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
XX CC coding sequences of the invention. Note: The sequence data for this
XX CC patent did not appear in the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 1946 BP; 423 A; 516 C; 519 G; 488 T; 0 U; 0 Other;

Query Match 79.1%; Score 17.4; DB 5; Length 1946;
Best Local Similarity 94.7%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 AAGTGATCGCTGCCGTGCT 22
DB 219 AAGTGATCGCTGCCGTGCT 201

RESULT 6
AAZ91119
ID AAZ91119 standard; DNA; 3661 BP.
XX AC AAZ91119;
XX DT 06-JUN-2000 (first entry)
XX DE Mutated Escherichia coli RNase E gene.
XX KW Ribonuclease E; RNase E; mutation; mRNA degradation; wild type;
XX KW protein truncation; recombinant gene expression; ds.
XX OS Escherichia coli.
XX OS Synthetic.
XX FH Key
XX FH Location/Qualifiers
XX FT 441..2198

```

```

FT      /*tag= a
FT      /product= "truncated mutant RNase E"
FT      replace(2196,G)
FT      /*tag= b
XX
XX      FR2782094-A1.
XX
XX      11-FEB-2000.
XX
XX      07-AUG-1998; 98FR-00010197.
XX
XX      07-AUG-1998; 98FR-00010197.
XX
XX      (CNRS ) CNRS CENT NAT RECH SCI.
XX
XX      Dreyfus M, Lopez P;
XX
XX      WPI; 2000-248411/22.
XX      P-PSDB; AAY80569.
XX
XX      New Escherichia coli mutants useful for producing recombinant proteins
XX      have mutation in ribonuclease E making it incapable of degrading
XX      messenger RNA.
XX
XX      Disclosure; Page: 34pp; French.
XX
XX      The invention relates to Escherichia coli strains in which the gene
XX      coding for ribonuclease E (RNase E) has a mutation such that the enzyme
XX      produced by expression of the gene is no longer capable of degrading
XX      messenger RNA (mRNA). This sequence represents the mutated RNase E gene
XX      which comprises a substitution of the G base at position 2196 of the wild
XX      type sequence with a T base (this sequence). The new codon created is a
XX      stop codon, resulting in a truncated protein having lost the C-terminal
XX      563 amino acids. The novel E. coli strains can be used to produce
XX      recombinant polypeptides by transforming them with a vector, especially a
XX      plasmid, containing a nucleotide sequence encoding one or more
XX      recombinant polypeptides and culturing the transformed cells for a time
XX      sufficient for expression of the polypeptide(s) in the cells. (Note: this
XX      sequence is not given in the specification but is generated from
XX      information given by the inventors)
XX
XX      SQ      Sequence 3661 BP; 930 A; 990 C; 1002 G; 739 T; 0 U; 0 Other;

Query Match      79.1%; Score 17.4; DB 3; Length 3661;
Best Local Similarity 94.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 AAGTGATCGCTGCCGTGGT 22
Db      3187 AAGTGATCGCTGCCGTGGT 3205

RESULT 7
AAZ91118
ID      AAZ91118 standard; DNA; 3661 BP.
XX
XX      AAZ91118;
XX
XX      06-JUN-2000 (first entry)
XX
XX      Escherichia coli RNase E gene.
XX
XX      Ribonuclease E; RNase E; mutation; mRNA degradation; wild type;
XX      protein truncation; recombinant gene expression; ds.
XX
XX      Escherichia coli.
XX
XX      Key      Location/Qualifiers
XX      CDS      441..3626
XX      /*tag= a
XX      /product= "RNase E"
XX
XX      FR2782094-A1.
XX
XX      11-FEB-2000.
XX
XX      07-AUG-1998; 98FR-00010197.
XX
XX      07-AUG-1998; 98FR-00010197.
XX
XX      (CNRS ) CNRS CENT NAT RECH SCI.
XX
XX      Dreyfus M, Lopez P;
XX
XX      WPI; 2000-248411/22.
XX      P-PSDB; AAY80569.
XX
XX      New Escherichia coli mutants useful for producing recombinant proteins
XX      have mutation in ribonuclease E making it incapable of degrading
XX      messenger RNA.
XX
XX      Disclosure; Page 18-23; 34pp; French.
XX
XX      The invention relates to Escherichia coli strains in which the gene
XX      coding for ribonuclease E (RNase E) has a mutation such that the enzyme
XX      produced by expression of the gene is no longer capable of degrading
XX      messenger RNA (mRNA). This sequence represents the wild type RNase E gene
XX      which is used as a basis for generating the mutant gene. The mutation is
XX      especially a substitution of the G base at position 2196 of this sequence
XX      with a T base (see AA291119). The new codon created is a stop codon,
XX      resulting in a truncated protein having lost the C-terminal 563 amino
XX      acids (AAY80569). The novel E. coli strains can be used to produce
XX      recombinant polypeptides by transforming them with a vector, especially a
XX      plasmid, containing a nucleotide sequence encoding one or more
XX      recombinant polypeptides and culturing the transformed cells for a time
XX      sufficient for expression of the polypeptide(s) in the cells
XX
XX      SQ      Sequence 3661 BP; 930 A; 990 C; 1003 G; 738 T; 0 U; 0 Other;

Query Match      79.1%; Score 17.4; DB 3; Length 3661;
Best Local Similarity 94.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 AAGTGATCGCTGCCGTGGT 22
Db      3187 AAGTGATCGCTGCCGTGGT 3205

RESULT 8
AAI24336/c
ID      AAI24336 standard; DNA; 412 BP.
XX
XX      AAI24336;
XX
XX      12-OCT-2001 (first entry)
XX
XX      Probe #14269 for gene expression analysis in human cervical cell sample.
XX
XX      Probe; human; microarray; gene expression; cervical epithelial cell;
XX      cervical cancer; ss.
XX
XX      Homo sapiens.
XX
XX      WO200157278-A2.
XX
XX      09-AUG-2001.
XX
XX      30-JAN-2001; 2001WO-US000670.
XX
XX      04-FEB-2000; 2000US-0180312P.
XX      26-MAY-2000; 2000US-0207456P.
XX      30-JUN-2000; 2000US-00608408.
XX      03-AUG-2000; 2000US-00632366.
XX      21-SEP-2000; 2000US-0234687P.
XX      27-SEP-2000; 2000US-0236359P.
XX      04-OCT-2000; 2000GB-00024263.

```

XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX DR WPI; 2001-488901/53.
 XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
 XX PT gene expression in human cervical epithelial cells.
 XX PS Claim 25; SEQ ID NO 14269; 487pp; English.
 XX CC The present invention relates to human single exon nucleic acid probes
 CC (SENp). The present sequence is one such probe. The SNPs are derived
 CC from human HeLa cells. The SNPs can be used to produce a single exon
 CC microarray, which can be used for measuring human gene expression in a
 CC sample derived from human cervical epithelial cells. By measuring gene
 CC expression, the probes are therefore useful in grading and/or staging of
 CC diseases of the cervix, notably cervical cancer. Note: The sequence data
 CC for this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 412 BP; 93 A; 102 C; 117 G; 100 T; 0 U; 0 Other;
 Query Match 78.2%; Score 17.2; DB 4; Length 412;
 Best Local Similarity 86.4%; Pred. No. 1.8e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 AGAAGTGTGCTCGCTCGGTGGT 22
 DB 168 AGAAGAGATCACTGCCCTGGT 147
 RESULT 9
 ABA69474/C
 ID ABA69474 standard; DNA; 412 BP.
 XX AC ABA69474;
 XX DT 01-FEB-2002 (first entry)
 XX DE Human foetal liver single exon nucleic acid probe #17779.
 XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 XX OS Homo sapiens.
 XX PN WO200157277-A2.
 XX PD 09-AUG-2001.
 XX PF 30-JAN-2001; 2001WO-US000669.
 XX PR 04-FEB-2000; 2000US-0180312P.
 XX PR 26-MAY-2000; 2000US-0207456P.
 XX PR 30-JUN-2000; 2000US-00608408.
 XX PR 03-AUG-2000; 2000US-00632366.
 XX PR 21-SEP-2000; 2000US-0234687P.
 XX PR 27-SEP-2000; 2000US-0236359P.
 XX PR 04-OCT-2000; 2000GB-00024263.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX DR WPI; 2001-483447/52.
 XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
 XX PT gene expression in human fetal liver.
 XX PS Claim 4; SEQ ID NO 17779; 639pp + Sequence Listing; English.
 XX CC The invention relates to a single exon nucleic acid probe for measuring
 CC human gene expression in a sample derived from human foetal liver. The
 CC single exon nucleic acid probes may be used for predicting, measuring and
 CC displaying gene expression in samples derived from human fetal liver. The
 CC present sequence is a single exon nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 412 BP; 93 A; 102 C; 117 G; 100 T; 0 U; 0 Other;
 Query Match 78.2%; Score 17.2; DB 4; Length 412;
 Best Local Similarity 86.4%; Pred. No. 1.8e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 AGAAGTGTGCTCGCTCGGTGGT 22
 DB 168 AGAAGAGATCACTGCCCTGGT 147
 RESULT 10
 AAI49606/C
 ID AAI49606 standard; DNA; 412 BP.
 XX AC AAI49606;
 XX DT 17-OCT-2001 (first entry)
 XX DE Probe #18292 used to measure gene expression in human placenta sample.
 XX KW Probe; microarray; human; placenta; antenatal diagnosis;
 XX KW genetic disorder; ss.
 XX OS Homo sapiens.
 XX PN WO200157272-A2.
 XX PD 09-AUG-2001.
 XX PF 30-JAN-2001; 2001WO-US000663.
 XX PR 04-FEB-2000; 2000US-0180312P.
 XX PR 26-MAY-2000; 2000US-0207456P.
 XX PR 30-JUN-2000; 2000US-00608408.
 XX PR 03-AUG-2000; 2000US-00632366.
 XX PR 21-SEP-2000; 2000US-0234687P.
 XX PR 27-SEP-2000; 2000US-0236359P.
 XX PR 04-OCT-2000; 2000GB-00024263.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX DR WPI; 2001-488897/53.
 XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
 XX PT gene expression in human placenta.
 XX PS Claim 25; SEQ ID NO 18292; 654pp; English.
 XX CC The present invention relates to single exon nucleic acid probes (SENp).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders
 XX SQ Sequence 412 BP; 93 A; 102 C; 117 G; 100 T; 0 U; 0 Other;
 Query Match 78.2%; Score 17.2; DB 4; Length 412;
 Best Local Similarity 86.4%; Pred. No. 1.8e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 AGAAGTGTGCTCGCTCGGTGGT 22

```
Db      168 AGAAAGAGATCACTGCCCTGGT 147
||||| |||| ||||| |||||
RESULT 11
ABA51446/c
ID ABA51446 standard; DNA; 412 BP.
XX
AC ABA51446;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human breast cell single exon nucleic acid probe #10141.
XX
KW Human; microarray; single exon probe; gene expression; breast; disease;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157271-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000662.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-496933/54.
XX
PT New spatially-addressable set of single exon nucleic acid probes, useful
PT for measuring gene expression in sample derived from human breast,
PT comprises number of single exon nucleic acid probes.
XX
PS Claim 4; SEQ ID NO 10141; 327pp + Sequence Listing; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting the
CC probes with a collection of detectably labelled nucleic acids derived
CC from mRNA of human breast, and then measuring the label bound to each
CC probe of the microarray. The probes are useful for verifying the
CC expression of regions of genomic DNA predicted to encode proteins. They
CC are useful for gene discovery, and for determining predisposition and/or
CC prognosing breast disease. Gene expression analysis is useful for
CC assessing the toxicity of chemical agents on cells. The microarray of
CC this invention presents a far greater diversity of probes for measuring
CC gene expression, with far less bias than expressed sequence tag
CC microarrays. The method is suitable for rapid production of functional
CC information from genomic sequence. The present sequence is a single exon
CC nucleic acid probe of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 412 BP; 93 A; 102 C; 117 G; 100 T; 0 U; 0 Other;
Query Match 78.2%; Score 17.2; DB 4; Length 412;
Best Local Similarity 86.4%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 AGAAAGTGATCGTCGCGTGGT 22
||||| ||||| ||||| |||||
Db      168 AGAAAGAGATCACTGCCCTGGT 147

RESULT 13
AAK43550/c
ID AAK43550 standard; DNA; 412 BP.
XX
AC AAK43550;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human breast cell single exon nucleic acid probe #10141.
XX
KW Human; microarray; single exon probe; gene expression; breast; disease;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157271-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000666.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488999/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts.
XX
PS Claim 4; SEQ ID NO 14865; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease. Note: The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 412 BP; 93 A; 102 C; 117 G; 100 T; 0 U; 0 Other;
Query Match 78.2%; Score 17.2; DB 4; Length 412;
Best Local Similarity 86.4%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 AGAAAGTGATCGTCGCGTGGT 22
||||| ||||| ||||| |||||
Db      168 AGAAAGAGATCACTGCCCTGGT 147

RESULT 13
AAK43550/c
ID AAK43550 standard; DNA; 412 BP.
XX
AC AAK43550;
XX
DT 06-NOV-2001 (first entry)
```



```
XX Human bone marrow expressed single exon probe SEQ ID NO: 18107.
DE
XX Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
KW
XX Homo sapiens.
OS
XX WO200157276-A2.
PN
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US000668.
PF
XX
XX 04-FEB-2000; 2000US-0180312P.
PR
XX 26-MAY-2000; 2000US-0207456P.
PR
XX 30-JUN-2000; 2000US-00608408.
PR
XX 03-AUG-2000; 2000US-00632366.
PR
XX 21-SEP-2000; 2000US-0234687P.
PR
XX 27-SEP-2000; 2000US-0236359P.
PR
XX 04-OCT-2000; 2000GB-00024263.
PR
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human bone marrow.
XX
XX Example 4; SEQ ID NO 18107; 658pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is one of
XX the probes of the invention
XX
XX Sequence 412 BP; 93 A; 102 C; 117 G; 100 T; 0 U; 0 Other;
XX
XX Query Match 78.2%; Score 17.2; DB 4; Length 412;
XX Best Local Similarity 86.4%; Pred. No. 1.8e+02;
XX Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX QY 1 AGAAGTGCCTGCTGCGTGGT 22
XX 168 AGAAGAGATCACTGCCCTGGT 147
XX
XX Db
XX
XX RESULT 14
XX AAK17729/c
XX ID AAK17729 standard; DNA; 412 BP.
XX
XX AC AAK17729;
XX
XX DT 05-NOV-2001 (first entry)
XX
XX Human brain expressed single exon probe SEQ ID NO: 17720.
XX
XX Human; brain expressed exon; gene expression analysis; probe; microarray;
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;
KW ss.
KW
XX Homo sapiens.
XX
XX WO200157275-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000667.
XX
```

```
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
PR (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains.
XX
XX Example 4; SEQ ID NO 17720; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is one of the probes of the
XX invention
XX
XX Sequence 412 BP; 93 A; 102 C; 117 G; 100 T; 0 U; 0 Other;
XX
XX Query Match 78.2%; Score 17.2; DB 4; Length 412;
XX Best Local Similarity 86.4%; Pred. No. 1.8e+02;
XX Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX QY 1 AGAAGTGCCTGCTGCGTGGT 22
XX 168 AGAAGAGATCACTGCCCTGGT 147
XX
XX Db
XX
XX RESULT 15
XX ABS43195/c
XX ID ABS43195 standard; DNA; 412 BP.
XX
XX AC ABS43195;
XX
XX DT 25-FEB-2003 (first entry)
XX
XX Human liver single exon probe, SEQ ID NO 18185.
XX
XX Human; single exon nucleic acid probe; liver; cirrhosis;
KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
KW coronary heart disease; ss.
KW
XX Homo sapiens.
XX
XX WO200157273-A2.
XX
XX PN
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US000664.
XX
XX 04-FEB-2000; 2000US-0180312P.
PR
XX 26-MAY-2000; 2000US-0207456P.
PR
XX 30-JUN-2000; 2000US-00608408.
PR
XX 03-AUG-2000; 2000US-00632366.
PR
XX 21-SEP-2000; 2000US-0234687P.
PR
XX 27-SEP-2000; 2000US-0236359P.
PR
XX 04-OCT-2000; 2000GB-00024263.
PR
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488998/53.
XX
```

XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human adult liver.
XX Claim 4; SEQ ID NO 18185; 658pp; English.
XX
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult liver.
CC (I) may be used for predicting, measuring and displaying gene expression
CC in samples derived from human adult liver. The genes identified may be
CC involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC associated with coronary heart disease. ABS25011-ABS51005 represent human
CC liver single exon nucleic acid probes of the invention. Note: The
CC sequence information for this patent does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 412 BP; 93 A; 102 C; 117 G; 100 T; 0 U; 0 Other;
Query Match 78.2%; Score 17.2; DB 4; Length 412;
Best Local Similarity 86.4%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 AGAAAGTGATCGCTGCCGTGGT 22
Db 168 AGAAGAGATCACTGCCCTGGT 147
||||| ||||| ||||| ||||| |||||
||||| ||||| ||||| ||||| |||||

Search completed: November 26, 2005, 13:40:50
Job time : 16.1112 secs

```
/db_xref="UniProt/TREMBL:O5T664"
/translations="MRNDSGPPPSVFNQNETFANI1KPTVVOQARIANGILGDF
I1HYDNRSGSDIGIIVGIFVHYFAPKDPPLPKNVVFLDSSASMGTKLRQTKD
ALFTIHLDRPQDRFIIIGFSNRKIKWDLHLSVTPDSIRDGKVIYHMSPTGTDIN
GALQAIRLNLKTVKHSIGLVSILVLTGKPTVGTHYTLKILNNTREARAGQVC
IPTIGLNDVDFLLKLELGLNGLRTRVHEEDAGSLGIFYDEIRTPLLSIRIDY
PSSVVQATKLPFPNPNSEI1IACKLVDRKDLHLVETVATSNKKFII1LKTDPVR
PQAKGVGTSPGPGDGGDTNHERLERSYLTKELLSSWLOSDDEPEKERLQRAQ
ALAVSRFLTPFTSMKLRGVPMDGLEEAHGSAMGPEPVQSVRGATQPGPLK
KPYOPRIKISKTSVDGPHFVDFPLSLRTVCFNIDGQGDILRLVSDHRDGSVTYNG
ELICGAPAPNGHKRQTYRTITLIPKPSRYLEITPSRVILDDGDLVLPNCNSV
VGSKGLVSNANVTYIQGSIAFVILHLYKAPFHLGFIYIANSEGLSNC
HGLGLQFNADARLTEDPAGPSNLTPLLLQVGEQPEAVLTVKGHVPVWVKRKIY
NGEQIDCWFPARNNAKLIDGEYKDYLAHFPTGMDTPGNVQGALKLAALQWQHE
GQ"
```

gene

```
join(complement(50428..50596),complement(39256..39300),
complement(25552..25715),complement(24379..24480),
complement(20853..21103),
complement(AL355374.14:81450..81619),
complement(AL355374.14:80319..80435),
complement(AL355374.14:50238..50406),
complement(AL355374.14:44092..44401),
complement(AL355374.14:40790..41349),
complement(AL355374.14:37479..37532),
complement(AL355374.14:34005..34121),
complement(AL355374.14:30367..30744),
complement(AL355374.14:23606..27721))
/locus_tag="RP11-676E7.1-001"
join(complement(50428..50596),complement(39256..39300),
complement(25552..25715),complement(24379..24480),
complement(20853..21103),
complement(AL355374.14:81450..81619),
complement(AL355374.14:80319..80435),
complement(AL355374.14:50238..50406),
complement(AL355374.14:44092..44401),
complement(AL355374.14:40790..41349),
complement(AL355374.14:37479..37532),
complement(AL355374.14:34005..34121),
complement(AL355374.14:30367..30744),
complement(AL355374.14:23606..27721))
/locus_tag="RP11-676E7.1-001"
```

mRNA

```
join(complement(50428..50596),complement(39256..39300),
complement(25552..25715),complement(24379..24480),
complement(20853..21103),
complement(AL355374.14:81450..81619),
complement(AL355374.14:80319..80435),
complement(AL355374.14:50238..50406),
complement(AL355374.14:44092..44401),
complement(AL355374.14:40790..41349),
complement(AL355374.14:37479..37532),
complement(AL355374.14:34005..34121),
complement(AL355374.14:30367..30744),
complement(AL355374.14:23606..27721))
/locus_tag="RP11-676E7.1-001"
```

```
/product="inter-alpha (globulin) inhibitor H5"
/notes="match: ESTs: AA134750.1 AA367892.1
AI089902.1 AI090076.1 AI339610.1 AI417871.1 AI079707.1
AI540791.1 AI543337.1 AI544674.1 AI544702.1 AI550125.1
AI550297.1 AI550915.1 AI554626.1 AI568595.1 AI569892.1
AI570737.1 AI570912.1 AI574207.1 AI574644.1 AI575373.1
AI577285.1 AM473748.1 BE893192.1 BE896223.1 BE898209.1
BF677721.1 BF689672.1 BF690566.1 BF726069.1 BF998687.1
BG006306.1 BG006608.1 BG282300.1 BG282310.1 BG326634.1
BG42562.1 BG435601.1 BG478666.1 BG676410.1 BG701677.1
BG767953.1 B1040702.1 B1553925.1 B1559648.1 BM016522.1
BM542130.1 BM542378.1 BM549311.1 BM557411.1 BM675641.1
BM906016.1 BM931606.1 BM985082.1 BM993373.1 BQ011450.1
BQ183379.1 BQ214877.1 BQ2425746.1 BQ574118.1 BQ722866.1
BQ80658.1 BU165852.1 BU171456.1 BU189399.1 CA307415.1
CA414285.1 H51540.1 N26400.1 N26407.1 R88167.1 T48709.1
W22615.1
match: CDNAs: AB075833.1 AK027375.1 AK075381.1 AL4833203.1"
join(complement(50428..50517),complement(39256..39300),
complement(25552..25715),complement(24379..24480),
complement(20853..21103),
complement(AL355374.14:81450..81619),
complement(AL355374.14:80319..80435),
complement(AL355374.14:50238..50406),
complement(AL355374.14:44092..44401),
complement(AL355374.14:40790..41349),
complement(AL355374.14:37479..37532),
complement(AL355374.14:34005..34121),
complement(AL355374.14:30367..30744),
complement(AL355374.14:27378..27721))
```

CDS

```
/gene="ITIH5"
/locus_tag="RP11-676E7.1-001"
/standard_name="OTTHUMP00000019062"
/notes="match: proteins: Q8TF48 Q96K70"
/codon_start=1
/product="inter-alpha (globulin) inhibitor H5"
/protein_id="Cali12955.1"
/db_xref="GI:55958060"
/db_xref="GOA:O5T665"
/db_xref="InterPro:IPR002035"
/db_xref="InterPro:IPR006587"
/db_xref="InterPro:IPR010600"
/db_xref="UniProt/TREMBL:O5T665"
/translation="MLLLGLCLGLSLCVSRSEAEQSGHSSQDGLRVPQVRLQQR
LTKPLMTFESVKSTIISRYAFTVTCRMLNRSAGEKTEI PRASAVIPSKDKAFLSLVE
GKPVQGEITEREKSGDRVSKRNKTEENGEKTEI PRASAVIPSKDKAFLSLVE
ELLQRRKIGKTESISVRPQQLSGRLSDVNIIESAGIASLEVLPLNSRQSGRGED
DSGPPSTVINQNETFANI1KPTVVOQARIANGILGDFIIRIDVNRSGSDIGIIV
NGYFVHYFAPKDPPLPKNVVFLDSSASMGTKLRQTKDALTFTIHLDRPQDRFII
IGFSNRKIKWDLHLSVTPDSIRDGKVIYHMSPTGTDINGALQRAIRLNLKTVKHS
IGRSVSLIVFLTLDGKPTVGETHILKILNNTREARAGQVCIFTIGLNDVDFLLKLE
SLNCGLTRVHEEDAGSLGIFYDEIRTPLLSIRIDYPPSSVVQATKLPFPNPNSEI1
GSEII1IAGKLVDRKDLHLVETVATSNKKFII1LKTDPVRPQAKGVGTSPGPGD
EGDTNHERLERSYLTKELLSSWLOSDDEPEKERLQRAQALAVSRFLTPFTSMKLR
GPVPRMDGLEEAHGSAMGPEPVQSVRGATQPGPLKPYOPRIKISKTSVDGPHF
HFVDPPLSLRTVCFNIDGQGDILRLVSDHRDGSVTYNGELIGAPAPNGHKRQTY
```

Query Match 100.0%; Score 26; DB 8; Length 174361;
Best Local Similarity 100.0%; Pred. No. 0.24; Indels 0; Gaps 0;
Matches 26; Conservative 0; Mismatches 0;

QY 1 GCGACACCAATTTGATGCTTTAAGA 26
DB 139649 GCGACACCAATTTGATGCTTTAAGA 139674

RESULT 8
AL391687 194563 bp DNA linear HTG 20-OCT-2001
LOCUS Homo sapiens chromosome 10 clone RP11-299D3, 2 unordered pieces.
DEFINITION AL391687
ACCESSION AL391687
VERSION HTG; HTGS PHASE1; HTGS_CANCELLED.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1
Almeida J.
Direct Submission
Submitted (16-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jul 12, 2001 this sequence version replaced gi:13990116.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA299D3
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 194262 bases at least Q40
Consensus quality: 194328 bases at least Q30
Consensus quality: 194372 bases at least Q20
Insert size: 194463; sum-of-contigs
Insert size: 173753; 41.5% error; agarose-fp
Quality coverage: 13.33x in Q20 bases; sum-of-contigs Quality
coverage: 14.92x in Q20 bases; agarose-fp

```

-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 172067: contig of 172067 bp in length
* 172068 172167: gap of 100 bp
* 172168 194563: contig of 22396 bp in length.
FEATURES             Location/Qualifiers
     source           1..194563
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone_lib="RPCI-11.2"
                     /clone="RP11-299D3"
misc_feature         1..172067
                     /notes="assembly fragment:00152
                     fragment_chain:1"
misc_feature         172168..194563
                     /notes="assembly fragment:03892
                     fragment_chain:1
                     clone_end:T7
                     vector_side:right"
ORIGIN
Query Match         100.0%; Score 26; DB 14; Length 194563;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCGAACACCAATTGATGCTTTAAGA 26
    |||||
Db 58795 GCGAACACCAATTGATGCTTTAAGA 58820

RESULT 9
CP000079.01
WPCOMMENT
Sequence split into 12 fragments LOCUS CP000079 Accession CP000079
Fragment Name      Begin      End
CP000079_00       1        110000
CP000079_01      100001    210000
CP000079_02      200001    310000
CP000079_03      300001    410000
CP000079_04      400001    510000
CP000079_05      500001    610000
CP000079_06      600001    710000
CP000079_07      700001    810000
CP000079_08      800001    910000
CP000079_09      900001   1010000
CP000079_10     1000001   1110000
CP000079_11     1100001   1130447
Continuation (2 of 12) of CP000079 from base 100001 (CP000079 Leishmania major strain F4

Query Match         77.7%; Score 20.2; DB 2; Length 110000;
Best Local Similarity 88.0%; Pred. No. 1.5e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 CGAACACCAATTGATGCTTTAAGA 26
    |||||
Db 87042 CGAACACCAATTGATGCTTTAAGA 87066

RESULT 10
AX003309/c
LOCUS
DEFINITION
Sequence 2 from Patent WO9929845.
ACCESSION
AX003309
VERSION
AX003309.1 GI:9927126

```

```

KEYWORDS
SOURCE             Mus sp.
ORGANISM           Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS           Angulo-Mora, J.F. and Mauffrey, P.
TITLE             Sequences coding for kin17 protein and their applications
JOURNAL           Patent: WO 9929845-A 2 17-JUN-1999;
                  ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE
                  (FR)
FEATURES           Location/Qualifiers
     source         1..1102
                     /organism="Mus sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:10095"
ORIGIN
Query Match        76.2%; Score 19.8; DB 6; Length 1102;
Best Local Similarity 91.3%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCGAACACCAATTGATGCTTTA 23
    |||||
Db 944 GCTAACACCAATTGATGCTTCA 922

RESULT 11
AX003331/c
LOCUS
DEFINITION
Sequence 24 from Patent WO9929845.
ACCESSION
AX003331
VERSION
AX003331.1 GI:9927146
KEYWORDS
SOURCE             Mus sp.
ORGANISM           Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS           Angulo-Mora, J.F. and Mauffrey, P.
TITLE             Sequences coding for kin17 protein and their applications
JOURNAL           Patent: WO 9929845-A 24 17-JUN-1999;
                  ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE
                  (FR)
FEATURES           Location/Qualifiers
     source         1..1390
                     /organism="Mus sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:10095"
ORIGIN
Query Match        76.2%; Score 19.8; DB 6; Length 1390;
Best Local Similarity 91.3%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCGAACACCAATTGATGCTTTA 23
    |||||
Db 1232 GCTAACACCAATTGATGCTTCA 1210

RESULT 12
MKKIN17/c
LOCUS
DEFINITION
Sequence 1414 bp mRNA for kin17 protein.
ACCESSION
X58472
VERSION
X58472.1 GI:1045209
KEYWORDS
Kin17 gene; kin17 protein; zinc-finger protein.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

```


Series: IRAK Plate: 123 Row: d Column: 16
This clone was selected for full length sequencing because it passed the following selection criteria: GenomScan gene prediction, Similarity but not identity to protein.

FEATURES
source
1..2710
/organism="Mus musculus"
/mol_type="mRNA"
/strain="mix FVB/N, C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:5026202"
/tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months old, gross tissue."
/clone_lib="NCI CGAP_Mam5"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

ORIGIN
Query Match 76.2%; Score 19.8; DB 9; Length 2710;
Best Local Similarity 91.3%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCGACACCAATTTGATGCTTTA 23
Db 2634 GCTAACCAATTTGATGCTTCA 2612

RESULT 14
CR9311708/c
LOCUS CR9311708 19113 bp DNA linear BCT 30-JUN-2005
DEFINITION Streptococcus pneumoniae strain 1095/39 (serotype 36).
ACCESSION CR9311708
VERSION CR9311708.1 GI:68644250
KEYWORDS
SOURCE Streptococcus pneumoniae
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

REFERENCE 1 (bases 1 to 19113)
AUTHORS Bentley,S.D., Aanensen,D., Mavroidi,A., Saunders,D., Rabinowitz,E., Collins,M., Donaghy,K., Harris,D., Kalltoft,M.S., Murphy,L., Quail,M.A., Samuel,G., Skovsted,I.C., Barrell,B.G., Reeves,P., Parkhill,J. and Spratt,B.G.
TITLE Genetic analysis of the capsular biosynthetic locus from all 90 serotypes of Streptococcus pneumoniae
JOURNAL Unpublished
AUTHORS Bentley,S.D.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-2004) Submitted on behalf of the Pathogen Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: sdb@sanger.ac.uk

COMMENT NOTE: This sequence was generated from a PCR product representing the region from dexB to aliA and is not necessarily responsible for the expressed capsule serotype. For a detailed description of how CDS products were predicted see the associated publication.
Location/Qualifiers
1..19113
/organism="Streptococcus pneumoniae"
/mol_type="genomic DNA"
/strain="1095/39"
/db_xref="taxon:1313"
<1..137
/genes="dexB"
/locus tag="SPC36_0001"
<1..137
/genes="dexB"
/locus tag="SPC36_0001"
/genes="dexB"
/locus tag="SPC36_0001"
/codon_start=3
/transl_table=11
/protein_id="CAI34356.1"
/db_xref="GI:68644251"
/translation="SNEEQDLTVGKVKSVLIENLAQEVFEKQLVPWDVFCVEMTD

FEATURES
source
1..19113
/organism="Streptococcus pneumoniae"
/mol_type="genomic DNA"
/strain="1095/39"
/db_xref="taxon:1313"
<1..137
/genes="dexB"
/locus tag="SPC36_0001"
<1..137
/genes="dexB"
/locus tag="SPC36_0001"
/codon_start=3
/transl_table=11
/protein_id="CAI34356.1"
/db_xref="GI:68644251"
/translation="SNEEQDLTVGKVKSVLIENLAQEVFEKQLVPWDVFCVEMTD

gene
complement(284..1659)
/gene="tnp"
/locus tag="SPC36_0002"
pseudo
complement(join(284..400,400..1659))
/gene="tnp"
/locus tag="SPC36_0002"
/note="member of homology group 0017 25"
pseudo
/codon_start=1
/transl_table=11
/product="putative IS1202 transposase"
complement(676..1176)
/gene="tnp"
/locus tag="SPC36_0002"
/note="HMPfam hit to PF00665, Integrase, catalytic region, score 5.4e-19"
pseudo
1835..3280
/gene="wzg"
/locus tag="SPC36_0003"
1835..3280
/gene="wzg"
/locus tag="SPC36_0003"
/note="member of homology group 0000 90"
/codon_start=1
/transl_table=11
/product="integral membrane regulatory protein Wzg"
/protein_id="CAI34358.1"
/db_xref="GI:68644252"
/translation="MSRRFKRSRSQKVRNIVLLTYLLVCFLLFLFKNLIAP
RYLNIVTALVLLVGLLIYKAEKFTILVFLSVLSVFLVQVQVGLTLNR
LNATNYSYISVAVLADSKNTVLTSTVPTGTONENIKLIADIKSSQNTDLT
VDSSSYLAAYKSLIAGTKAIVLNSVFENISEYSPDYASKIKLYTKGFKKVEAP
KTSKQSNLYVSGIDTGPITSSVSRSQVNIIMTVNRDTKILLTTTPRDVYVPIADG
GNKQDKLTAGIYGVDSIHTLENLYGVDIYVYRLNFTSFLKMLDILGGVDVNDQ
EFSALHGKFPFVGNVHLDSEQALGFVRERYSLADRRGRNQKVIIVAIKQLTST
EVLKNYSILQGLDQSLQNMPIETMIDLVNTQLESNGGKYNVSQDLKGTGRMDLPSY
AMPDNLVYMEVDDSSLAIVKAAIQDVMGR"
1835..1951
/gene="wzg"
/locus tag="SPC36_0003"
/note="Signal peptide predicted for SPC1724 by SignalP 2.0
HMM (Signal peptide probability 0.999) with cleavage site
probability 0.540 between residues 39 and 40"
join(1883..1942,1970..2038,2057..2125)
/gene="wzg"
/locus tag="SPC36_0003"
/note="3 probable transmembrane helices predicted for
SPC1724 by TMHMM2.0 at aa 17-36, 46-68 and 75-97"
2042..2386
/gene="wzg"
/locus tag="SPC36_0003"
/note="HMPfam hit to PF02916, DNA polymerase processivity
factor, score 2.7e-62"
2564..3007
/gene="wzg"
/locus tag="SPC36_0003"
/note="HMPfam hit to PF03816, Cell envelope-related
transcriptional attenuator, score 3.2e-70"
3282..4013
/gene="wzh"
/locus tag="SPC36_0004"
3282..4013
/gene="wzh"
/locus tag="SPC36_0004"
/EC_number="3.1.3.48"
/note="member of homology group 0001 90"
/codon_start=1
/transl_table=11
/product="protein-tyrosine phosphatase Wzh"
/protein_id="CAI34359.1"

```

/db_xref="GI:68644253"
/translation="MIDHSHIVPDVDDGPKSREESKALLABSVROGVRTIVSTSHR
KMPETPEEKIAFLQVREIAKEAVDDLVVAGAEYYTLDALEKLEKEIETLND
RYALIEFSEMTSYREIHTGJNLIUMGITTIVIAHIERYDALENNEKRVCSLIDMGCT
QINSYHLKPKFGEKTKPKKQARARYFLERDLVHVASDMHNLDSRPPHMAEAYDLVT
QKYGAAQQLFVDNPKRIIMDQLI"
misc_feature
3603..3908
/feature="wzh"
/locus_tag="SPC36_0004"
/note="HMPFam hit to pf02811, PHP, C-terminal, score
1.9e-10"
gene
4022..4714
/feature="wzd"
/locus_tag="SPC36_0005"
CDS
4022..4714
/feature="wzd"
/locus_tag="SPC36_0005"
/note="member of homology group 0002 90"
/codon_start=1
/transl_table=11
/product="capsular polysaccharide biosynthesis protein
wzd"
/protein_id="CAI34360.1"
/db_xref="GI:68644254"
/translation="MKQNTLEIDVLQFLMKRKLVLVAILTSSVAFYSTFVI
KPEFTSTRIVVRNQEESGLNQDLQAGSYLVKDYREIILSQDYLEVISDLKLD
LTPKGLANKIKVTPVDTRIVSVNDKRVPEEASRIANSREVAQKIISITRVSQDT
LLEAPPAISLSLNIKRNTLIGLAGIGTSVIFLELLDTHVKKRPEDIEDTLQMT
LLGVVPLNSLK"
4040..4456
/feature="wzd"
/locus_tag="SPC36_0005"
/note="HMPFam hit to pf02706, Lipopolysaccharide
biosynthesis, score 1.2e-53"
join(4088..4153,4556..4624)
/feature="wzd"
/locus_tag="SPC36_0005"
/note="2 probable transmembrane helices predicted for
SPC1726 by TMHMM2.0 at aa 23-44 and 179-201"
4562..4639
/feature="wzd"
/locus_tag="SPC36_0005"
/note="PS00217 Sugar transport proteins signature 2."
4724..5407
/feature="wze"
/locus_tag="SPC36_0006"
4724..5407
/feature="wze"
/locus_tag="SPC36_0006"
/EC_number="2.7.1.112"
/note="member of homology group 0003 90"
/codon_start=1
/transl_table=11
/product="tyrosine-protein kinase Wze"
/protein_id="CAI34361.1"
/db_xref="GI:68644255"
/translation="MPTLEIAQKLEFIKKAEEYNALCTNIQLSGDKLVISVTSVN
PGEGLTTSVNIARISFARAGYKLLIDGTRNSVISGFKSREKITGLTEFLSGTADL
SHGLCDTNIENLFTQSGSVSPNTALIQSKNFNDMIETLRKYFAYIIVDTAPIGIVI
DAAIITQKCDASILVTATGEVNRKRDVQAKQOLETKGLFLGVNLKLDISVDKYGVI
GSYGNYGKK"
5423..6790
/feature="wcha"
/locus_tag="SPC36_0007"
5423..6790
/feature="wcha"
/locus_tag="SPC36_0007"
/note="member of homology group 0005 66"
/codon_start=1
/transl_table=11
/product="undecaprenylphosphate glucosylphosphotransferase
wcha (initial sugar transferase)"
/protein_id="CAI34362.1"

```

```

/db_xref="GI:68644256"
/translation="MDEKGLKIFLAVLQSIIVILLVYFLSFVRETELEERSMWILYLL
HPEVYFSSVGNFPRKGYLVBFNSTIRVIFFAIAISVLPFIAERFSISRRGMVYF
LITGLSLLYLANPLVKYKWHVFENLKNKSKILLTLVTKNMKVLKLLSDELSSWKL
VAVSVDKSDFOHQHDKIPVIEKELIEFAHEVDEVFVNLPGESYDIGEIIISRFETMG
IDVTYNLAKPDRKLNKQKIHENEGLNVVFTSNFYKTSIKRSLDLCGATIGLVL
FAIASLVPLRKDGKGPAPFAQTRIGKGRHFTFYKFRSMRIDAISAKELQMDQNTM
QGMGFMDNDPRITKIGRITKTSDELQFQYVNLKGMSLVGRMTRPTVEDEYHYTPE
OKERLSFKPGITGLWQVSGRSKIIDFDDVVKLDVAVIDNWTIWKDIEILLTKTVKVFEM
RDGAK"
Query Match 76.2%; Score 19.8; DB 1; Length 19113;
Best Local Similarity 91.3%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 3 GAACACCAATTGCTGCTTAAG 25
||||| ||||| ||||| ||||| |||||
Db 15957 GAACACGAATTTTATGCTTTAAG 15935

RESULT 15
AP008173
LOCUS
DEFINITION
Lotus corniculatus var. japonicus chromosome 1 clone Lj725J06, ***
SEQUENCING IN PROGRESS ***, 11 unordered pieces.
ACCESSION
AP008173.1 GI:56806479
VERSION
HTG; HTGS_PHASE1.
KEYWORDS
Lotus corniculatus var. japonicus (Lotus japonicus)
SOURCE
Lotus corniculatus var. japonicus
ORGANISM
Lotus.
1
REFERENCE
1
Kaneko, T., Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
Structural Analysis of a Lotus japonicus Genome. XI. Sequence
Features and Mapping of Nine hundred twenty-one TAC Clones
Unpublished
2 (bases 1 to 86263)
Sato, S.
Direct Submission
Submitted (26-OCT-2004) Shusei Sato, Kazusa DNA Research Institute,
Department of Plant Gene Research; 2-6-7 Kazusa-kamatarai, Kisarazu,
Chiba, 252-0818, Japan (E-mail: ssato@kazusa.or.jp).
URL: http://www.kazusa.or.jp/, Tel: 81-438-52-3335 (ex. 2337),
Fax: 81-438-52-3934)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 11721: contig of 11721 bp in length
* 11722 11821: gap of unknown length
* 11822 12921: contig of 1100 bp in length
* 12922 13021: gap of unknown length
* 13022 16157: contig of 3136 bp in length
* 16158 16257: gap of unknown length
* 16258 19596: contig of 3339 bp in length
* 19597 19696: gap of unknown length
* 19697 23700: contig of 4004 bp in length
* 23701 23800: gap of unknown length
* 23801 29711: contig of 5911 bp in length
* 29712 29811: gap of unknown length
* 29812 36936: contig of 7125 bp in length
* 36937 37036: gap of unknown length
* 37037 48212: contig of 11176 bp in length
* 48213 48312: gap of unknown length
* 48313 61092: contig of 12780 bp in length
* 61093 61192: gap of unknown length

```



```

* 61193 77610: contig of 16418 bp in length
* 77611 77710: gap of unknown length
* 77711 86263: contig of 8553 bp in length.
FEATURES
    source
        1..86263
            /organism="Lotus corniculatus var. japonicus"
            /mol_type="genomic DNA"
            /variety="japonicus"
            /db_xref="taxon:34305"
            /chromosome="1"
            /clone="LjT25J06"
            /clone_lib="LjT library"
            /notes="TAC clone: TM1595, synonym: Lotus japonicus"
        11722..11821
            /estimated_length=unknown
        12922..13021
            /estimated_length=unknown
        16158..16257
            /estimated_length=unknown
        19597..19696
            /estimated_length=unknown
        23701..23800
            /estimated_length=unknown
        29712..29811
            /estimated_length=unknown
        36937..37036
            /estimated_length=unknown
        48213..48312
            /estimated_length=unknown
        61093..61192
            /estimated_length=unknown
        77611..77710
            /estimated_length=unknown
    ORIGIN

```

```

Query Match      76.2%; Score 19.8; DB 14; Length 86263;
Best Local Similarity 91.3%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      4 AACACCAATTGATGCTTTAAGA 26
      ||||| ||||| ||||| ||||| |||||
DB     10506 AACACGAATTGATGCTTCAGA 10528

```

Search completed: November 26, 2005, 19:11:07
 Job time : 141.192 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 12:48:57 ; Search time 16.6769 Seconds
(without alignments)
10390.557 Million cell updates/sec

Title: US-09-555-529-19
Perfect score: 26
Sequence: 1 gcgaacacaaattgatgctttaaga 26

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21.*
1: geneseqn1980s.*
2: geneseqn1990s.*
3: geneseqn2000s.*
4: geneseqn2001as.*
5: geneseqn2001bs.*
6: geneseqn2002as.*
7: geneseqn2002bs.*
8: geneseqn2003as.*
9: geneseqn2003bs.*
10: geneseqn2003cs.*
11: geneseqn2003ds.*
12: geneseqn2004as.*
13: geneseqn2004bs.*
14: geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	26	100.0	26	AAX85567	AAX85567 PCR prime
C 2	26	100.0	591	AAX34095	AAX34095 Human col
C 3	26	100.0	1002	AAX85551	AAX85551 CDNA of a
C 4	26	100.0	1296	AAX85549	AAX85549 cDNA of a
5	22	84.6	398	ACF87901	ACF87901 Human SIR
C 6	21.2	81.5	679	ADQ56943	ADQ56943 Novel can
C 7	19.8	76.2	1102	AAX85550	AAX85550 CDNA of a
C 8	19.8	76.2	1390	AAX85570	AAX85570 CDNA of a
C 9	19.8	76.2	1458	AAQ79936	AAQ79936 Murine Ki
10	19.6	75.4	915	AAQ79762	AAQ79762 Helicobac
11	19.6	75.4	1024	ABX66121	ABX66121 Helicobac
12	19.6	75.4	1116	AAX14506	AAX14506 H. pylori
C 13	18.8	72.3	4328	AAX29635	AAX29635 Arabidops
C 14	18.8	72.3	4505	AAV15448	AAV15448 Arabidops
C 15	18.8	72.3	14113	AAA39651	AAA39651 Fusion co
C 16	18.8	72.3	14113	AAA90096	AAA90096 PR-1 prom
C 17	18.8	72.3	14113	AAH25853	AAH25853 PR-1 prom
C 18	18.8	72.3	134499	AAX22286	AAX22286 BAC conta
19	18.2	70.0	2253	AAD54900	AAD54900 Arabidops

20	18	69.2	217	3	AAC30472	AAC30472 Human sec
21	18	69.2	407	10	ADF81241	ADF81241 Leukaemia
22	18	69.2	466	10	ADF82611	ADF82611 Leukaemia
23	18	69.2	516	13	ACN56786	ACN56786 Cotton gy
24	18	69.2	528	13	ACN62182	ACN62182 Cotton gy
C 25	18	69.2	559	13	ACN60657	ACN60657 Cotton gy
C 26	18	69.2	559	13	ACN51102	ACN51102 Cotton an
C 27	18	69.2	574	13	ACN56699	ACN56699 Cotton gy
C 28	18	69.2	599	13	ACN62609	ACN62609 Cotton de
C 29	18	69.2	606	13	ACN62525	ACN62525 Cotton de
30	18	69.2	1839	10	ADF09893	ADF09893 Human pap
31	18	69.2	4498	13	ADX64079	ADX64079 Plant ful
32	18	69.2	110000	13	ABD32909	ABD32909 (7 of
C 33	18	69.2	158811	12	ADQ19901	ADQ19901 Human sof
34	17.6	67.7	309	8	ACF72766	ACF72766 Staphyloc
35	17.6	67.7	561	11	ACL30673	ACL30673 Rice abio
36	17.6	67.7	672	10	ADF01210	ADF01210 Bacterial
C 37	17.6	67.7	697	6	ABK34383	ABK34383 Human CDN
C 38	17.6	67.7	837	8	ACA30257	ACA30257 Prokaryot
C 39	17.6	67.7	943	6	ABV77378	ABV77378 Glucose f
C 40	17.6	67.7	1024	10	ADC86948	ADC86948 Human GPC
C 41	17.6	67.7	1056	8	ACA38525	ACA38525 Prokaryot
42	17.6	67.7	1200	10	ACC60903	ACC60903 Gene sequ
43	17.6	67.7	1200	10	ADK62329	ADK62329 Disease c
44	17.6	67.7	1238	13	ADT18265	ADT18265 Plant cDN
C 45	17.6	67.7	1247	6	ABL40378	ABL40378 Human pho

ALIGNMENTS

RESULT 1

AAX85567
ID AAX85567 standard; cDNA; 26 BP.

AC AAX85567;
XX

XX 07-SEP-1999 (first entry)
XX

XX PCR primer and probe used to detect human kin17 nucleic acids.
XX

DE Human; kin17 protein; cell proliferation; fertility; probe;
XX hyperproliferative disease; protein interaction; curved DNA;

KW HIV replication; HIV integration; repair enzyme; PCR primer; as.
KW

XX Synthetic.
OS

OS Homo sapiens.
OS

XX FR2772046-A1.
PN

XX 11-JUN-1999.
PD

XX 09-DEC-1997; 97FR-00015536.
PF

XX 09-DEC-1997; 97FR-00015536.
PR

XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
PA

XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MUF;
PI

XX WPI; 1999-359999/31.
DR

XX New DNA coding for human kin17 protein - useful for controlling cell
XX proliferation or fertility.
PT

XX Claim 7; Page 5; 69pp; French.
PS

CC AAX85553-69 represent PCR primers and probes used to detect nucleic acids
CC encoding human kin17 protein. The mammalian kin17 protein is useful for
CC preparing a medicament for controlling cell proliferation or for
CC controlling fertility. The medicaments can also be used to treat
CC hyperproliferative diseases. Fragments between amino acids 55 and 235
CC (preferably between amino acids 129 and 228) of a mammalian kin17 protein

CC are useful for regulating the interaction between proteins and curved
CC DNA. The fragment can be used to block replication of HIV or its
CC integration into the human genome or to target repair enzymes to curved
CC DNA sites. Expression vectors for kin17 can be used for controlling cell
CC proliferation

XX SQ Sequence 26 BP; 9 A; 5 C; 5 G; 7 T; 0 U; 0 Other;

Query Match 100.0%; Score 26; DB 2; Length 26;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGAACACCAATTGATGCTTTAAGA 26
Db 1 GCGAACACCAATTGATGCTTTAAGA 26

RESULT 2

AAH34095/c
ID AAH34095 standard; cDNA; 591 BP.

XX AC AAH34095;

XX DT 03-SEP-2001 (first entry)

XX DE Human colon cancer antigen encoding cDNA SEQ ID NO:1177.

XX KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX KW colorectal carcinoma; ss.

XX OS Homo sapiens.

XX PN WO200122920-A2.

XX XX 05-APR-2001.

XX PF 28-SEP-2000; 2000WO-US026524.

XX PR 29-SEP-1999; 99US-0157137P.

XX PR 03-NOV-1999; 99US-0163280P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Ruben SM, Barash SC, Birse CE, Rosen CA;

XX WPI; 2001-235357/24.

XX DR P-PSDB; AAG74690.

XX PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers.

XX PS Claim 1; Page 2977-2978; 9803pp; English.

XX CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where the
XX proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene therapy
XX and vaccine production. N and P may be used in the prevention, diagnosis
XX and treatment of diseases associated with inappropriate P expression. For
XX example, N and P may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of P by expressing inactive proteins or to
XX supplement the patient's own production of P. Additionally, N may be used
XX to produce the colon cancer-associated Ps, by inserting the nucleic acids
XX into a host cell and culturing the cell to express the proteins. N and P
XX can be used in the prevention, diagnosis and treatment of colorectal
XX carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent
XX sequences used in the exemplification of the present invention. N.B.
XX Pages 666 to 682 and page 7053 of the sequence listing were missing at
XX time of publication, meaning no sequences are present for SEQ ID NO:1027
XX to 1052, 7921 and 7922

XX SQ Sequence 591 BP; 199 A; 95 C; 114 G; 182 T; 0 U; 1 Other;

Query Match 100.0%; Score 26; DB 4; Length 591;
Best Local Similarity 100.0%; Pred. No. 0.057;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGAACACCAATTGATGCTTTAAGA 26
Db 326 GCGAACACCAATTGATGCTTTAAGA 301

RESULT 3

AAH85551/c
ID AAX85551 standard; cDNA; 1002 BP.

XX AC AAX85551;

XX DT 07-SEP-1999 (first entry)

XX DE cDNA of a gene coding for a truncated human kin17 protein.

XX KW kin17 protein; cell proliferation; fertility; hyperproliferative disease;
XX KW protein interaction; curved DNA; HIV replication; HIV integration;
XX KW repair enzyme; ss.

XX OS Homo sapiens.

XX PN FR2772046-A1.

XX PD 11-JUN-1999.

XX PF 09-DEC-1997; 97FR-00015536.

XX PR 09-DEC-1997; 97FR-00015536.

XX PA (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

XX PI Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;

XX WPI; 1999-359999/31.

XX DR New DNA coding for human kin17 protein - useful for controlling cell
XX proliferation or fertility.

XX PS Claim 5; Page 32; 69pp; French.

XX CC The present sequence encodes a truncated human kin17 protein with amino
XX acids 129-228 deleted. The mammalian kin17 protein is useful for
XX preparing a medicament for controlling cell proliferation or for
XX controlling fertility. The medicaments can also be used to treat
XX hyperproliferative diseases. Fragments between amino acids 55 and 235
XX (preferably between amino acids 129 and 228) of a mammalian kin17 protein
XX are useful for regulating the interaction between proteins and curved
XX DNA. The fragment can be used to block replication of HIV or its
XX integration into the human genome or to target repair enzymes to curved
XX DNA sites. Expression vectors for kin17 can be used for controlling cell
XX proliferation

XX SQ Sequence 1002 BP; 343 A; 189 C; 230 G; 240 T; 0 U; 0 Other;

Query Match 100.0%; Score 26; DB 2; Length 1002;
Best Local Similarity 100.0%; Pred. No. 0.062;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGAACACCAATTGATGCTTTAAGA 26
Db 1002 GCGAACACCAATTGATGCTTTAAGA 977

RESULT 4

AAH85549/c
ID AAX85549 standard; cDNA; 1296 BP.

XX AC AAX85549;


```

XX The present sequence encodes a murine kin17 protein. The mammalian kin17
CC protein is useful for preparing a medicament for controlling cell
CC proliferation or for controlling fertility. The medicaments can also be
CC used to treat hyperproliferative diseases. Fragments between amino acids
CC 55 and 235 (preferably between amino acids 129 and 228) of a mammalian
CC kin17 protein are useful for regulating the interaction between proteins
CC and curved DNA. The fragment can be used to block replication of HIV or
CC its integration into the human genome or to target repair enzymes to
CC curved DNA sites. Expression vectors for kin17 can be used for
CC controlling cell proliferation
XX
SQ Sequence 1390 BP; 482 A; 256 C; 353 G; 299 T; 0 U; 0 Other;
Query Match 76.2%; Score 19.8; DB 2; Length 1390;
Best Local Similarity 91.3%; Pred. No. 51;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GCGAACACCAATTGTGATGCTTTA 23
Db 1232 GCTAACACCAATTGTGATGCTTCA 1210
RESULT 9
AAQ79936/c
ID AAQ79936 standard; cDNA; 1458 BP.
XX
AC AAQ79936;
XX
XX 25-MAR-2003 (revised)
DT 06-SEP-1995 (first entry)
XX
XX Murine Kin17 cDNA.
DE
XX chromosomal rearrangement; kin17 protein; SOS DNA repair system; RecA;
KW genotoxic agent; zinc finger; DNA binding protein; ss.
KW
XX Mus musculus.
OS
XX
FH Key Location/Qualifiers
FT primer_bind complement(1..21)
FT /tag= b
FT /note= "Oligo L (AAQ79946) binds to complementary strand"
FT misc_feature 22..1434
FT /tag= 1
FT /label= kin17 cDNA
FT /note= "nucleotides 1-1414; the genomic DNA contains at
FT least two introns within this sequence, see Comments"
FT primer_bind 32..49
FT /tag= c
FT /note= "Oligo C (AAQ79938) binding site"
FT CDS 46..1221
FT /tag= a
FT /product= "Kin17"
FT /note= "N's in the sequence denote illegible residues"
FT primer_bind complement(67..86)
FT /tag= d
FT /note= "Oligo S (AAQ79947) binds to complementary strand"
FT primer_bind 274..297
FT /tag= e
FT /note= "Oligo D (AAQ79939) binding site"
FT primer_bind complement(339..360)
FT /tag= f
FT /note= "Oligo K (AAQ79945) binds to complementary strand"
FT primer_bind 451..474
FT /tag= g
FT /note= "Oligo J (AAQ79944) binding site"
FT primer_bind complement(550..567)
FT /tag= h
FT /note= "Oligo E (AAQ79940) binds to complementary strand"
FT primer_bind 802..825
FT /tag= i
FT /note= "Oligo F (AAQ79941) binding site"
FT

```

```

FT primer_bind complement(839..862)
FT /tag= j
FT /note= "Oligo G (AAQ79942) binds to complementary strand"
FT primer_bind complement(1435..1458)
FT /tag= k
FT /note= "Oligo B (AAQ79937) binds to complementary strand"
XX
XX PR2706487-A1.
XX
XX 23-DEC-1994.
XX
XX 15-JUN-1993; 93FR-00007171.
XX
XX 15-JUN-1993; 93FR-00007171.
XX
XX (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.
XX
XX Angulo-Mora JF, Tlissier A, Frelat G, Mauffrey P, Guilly M;
XX WPI; 1995-039031/06.
XX
XX Purified murine kin17 protein prepn. for detecting chromosomal
XX rearrangements - also related antibodies, human and murine DNA, primers,
XX probes and vectors, used to assess damage caused by genotoxic agents.
XX
XX Claim 9; Page 33; 54pp; French.
XX
XX The murine Kin17 protein includes a zinc finger domain (see AAR66766),
XX recognises single- and double-stranded DNA (partic. regions of secondary
XX structure), has apparent mol. wt. 43 kD and is recognised by both anti-
XX kin17 antibodies and antibodies against the RecA protein of E.coli. The
XX Kin17 protein is involved in DNA repair; it can be used to monitor
XX chromosomal rearrangements following exposure to genotoxic agents. The
XX kin17 cDNA sequence AAQ79936 consists of a 1414 nucleotide sequence,
XX flanked by primer binding sites; the genomic kin17 DNA is claimed in
XX which an intron is inserted at position 137 of the 1414 nucleotide cDNA
XX and a second intron is located between nucleotides 339-429 of the cDNA.
XX (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 1458 BP; 496 A; 272 C; 372 G; 311 T; 0 U; 7 Other;
Query Match 76.2%; Score 19.8; DB 2; Length 1458;
Best Local Similarity 91.3%; Pred. No. 51;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GCGAACACCAATTGTGATGCTTTA 23
Db 1277 GCTAACACCAATTGTGATGCTTCA 1255
RESULT 10
AAC97262
ID AAC97262 standard; DNA; 915 BP.
XX
XX AAC97262;
XX
XX 23-FEB-2001 (first entry)
XX
XX Helicobacter pylori bait polypeptide nucleotide sequence #34.
XX
XX Helicobacter pylori; two-hybrid system; protein-protein interaction;
XX antibacterial; bait polypeptide; gastric ulcer; ds.
XX
XX Helicobacter pylori.
XX
XX WO200066722-A1.
XX
XX 09-NOV-2000.
XX
XX 14-APR-2000; 2000WO-IB000603.
XX
XX 30-APR-1999; 99EP-00401066.
XX

```

PA (HYBR-) HYBRIGENICS SA.
 XX Legrain P, Selig L, Rain J;
 XX WPI: 2000-687535/67.
 DR P-PSDB; AAB52516.
 XX A two-hybrid system for identifying compounds useful in the treatment of
 PT e.g. gastric ulcers comprises producing a collection of recombinant cell
 PT clones.
 XX
 XX Example 5; Page 118-119; 267pp; English.
 XX
 CC The present sequence encodes a bait polypeptide used in a Helicobacter
 CC pylori two-hybrid screen to identify protein-protein interactions. The
 CC method is used to identify a recombinant cell clone expressing a prey
 CC polypeptide which is capable of interacting with the bait polypeptide.
 CC The two hybrid system is useful for screening compounds for antibacterial
 CC activity. It may be used in the treatment of gastric ulcers. The
 CC polynucleotides are useful as amplification primers or specific detection
 CC probes. The polypeptides, vectors or host cells can be used as immunogens
 CC to produce mono- or polyclonal antibodies. The polynucleotides,
 CC polypeptides, antibodies, vectors, host cells or modulating agents can be
 CC used to produce a pharmaceutical composition
 XX
 SQ Sequence 915 BP; 291 A; 161 C; 216 G; 247 T; 0 U; 0 Other;
 Query Match 75.4%; Score 19.6; DB 3; Length 915;
 Best Local Similarity 84.6%; Pred. No. 59;
 Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 GCGAACACCAATTGATGCTTTAAGA 26
 DB 536 GCGATCTCAAAATTGATGCTTTAAGA 561
 RESULT 11
 ABX66121
 ID ABX66121 standard; DNA; 1024 BP.
 XX
 AC ABX66121;
 XX
 DT 07-MAY-2003 (first entry)
 XX
 DE Helicobacter pylori selected interacting domain (SID) DNA #720.
 XX
 KW Protein-protein interaction; ulcer; selected interacting domain; SID;
 KW gene, ds.
 XX
 OS Helicobacter pylori.
 XX
 PN WQ200266501-A2.
 XX
 PD 29-AUG-2002.
 XX
 PF 28-DEC-2001; 2001WO-EP015428.
 XX
 XX 02-JAN-2001; 2001US-0259302P.
 XX
 PA (HYBR-) HYBRIGENICS.
 PA (INSP) INST PASTEUR.
 XX
 PI Legrain P, Rain J, Colland F, De Reuse H, Labigne A;
 DR WPI: 2002-674910/72.
 DR P-PSDB; AB051377.
 XX
 XX New complexes of protein-protein interactions in Helicobacter pylori,
 PT useful for identifying modulating compounds for treating or preventing
 PT ulcers in mammals.
 XX
 PS Claim 7; Page 260; 642pp; English.
 XX

CC The invention describes a complex of protein-protein interactions in
 CC Helicobacter pylori selected from 421 complexes given in the
 CC specification. The complex of protein-protein interactions are useful for
 CC screening for agents which modulate the interaction of proteins.
 CC Modulating compounds which binds to a targeted bacterial protein may be
 CC used for treating or preventing ulcers in a human or animal. This
 CC sequence encodes a selected interacting domain (SID), identified via
 CC protein-protein interactions
 XX
 SQ Sequence 1024 BP; 326 A; 183 C; 233 G; 282 T; 0 U; 0 Other;
 Query Match 75.4%; Score 19.6; DB 6; Length 1024;
 Best Local Similarity 84.6%; Pred. No. 60;
 Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 GCGAACACCAATTGATGCTTTAAGA 26
 DB 537 GCGATCTCAAAATTGATGCTTTAAGA 562
 RESULT 12
 AAX14506
 ID AAX14506 standard; DNA; 1116 BP.
 XX
 AC AAX14506;
 XX
 DT 31-MAR-1999 (first entry)
 XX
 DE H. pylori GHPO 1220 gene.
 XX
 KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
 KW peptic ulcer disease; ss.
 XX
 OS Helicobacter pylori.
 XX
 FH Key Location/Qualifiers
 FT CDS 70..1068
 FT /*tag= a
 XX
 PN WQ9843478-A1.
 XX
 PD 08-OCT-1998.
 XX
 PF 01-APR-1998; 98WO-US006371.
 XX
 PR 01-APR-1997; 97US-00833457.
 PR 24-JUN-1997; 97US-00881227.
 PR 29-JUL-1997; 97US-00902615.
 XX
 PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Kleanthous H, Al-Garawi A, Miller C, Tomb J, Oomen RP;
 XX
 DR WPI: 1998-542293/46.
 DR P-PSDB; AAW98787.
 XX
 XX New isolated Helicobacter polynucleotides - used to develop products for
 PT the diagnosis, prevention and treatment of Helicobacter infections and
 PT gastrointestinal diseases.
 PT
 XX Claim 1; Page 1713-1714; 2054pp; English.
 XX
 CC This sequence represents a polynucleotide of the invention. It was
 CC isolated from Helicobacter pylori and encodes a H.pylori GHPO protein.
 CC The polypeptides can be used for preventing or treating Helicobacter
 CC infections, and gastroduodenal diseases associated with these infections,
 CC including acute, chronic, and atrophic gastritis, and peptic ulcer for the
 CC diseases, e.g. gastric and duodenal ulcers. They can also be used for the
 CC production of antibodies. The products can also be used for detection and
 CC diagnosis
 XX
 SQ Sequence 1116 BP; 356 A; 203 C; 255 G; 302 T; 0 U; 0 Other;

Query Match 75.4%; Score 19.6; DB 2; Length 1116;
Best Local Similarity 84.6%; Pred. No. 61;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCGAACACCAATTGATGCTTTAAGA 26
|||||
Db 686 GCGATCTCAAAATTGATGCTTTAAGA 711

RESULT 13
AAD29635/c
ID AAD29635 standard; DNA; 4328 BP.

XX AC AAD29635;

XX DT 17-MAY-2002 (first entry)

XX DE Arabidopsis thaliana isochorismate synthase related DNA #1.

XX KW Isochorismate synthase; ICS; systemic acquired resistance; SAR; SA;
XX KW salicylic acid; disease resistance; powdery mildew; pathogenic nematode;
XX KW tobacco mosaic virus; TMV; tobacco necrosis virus; TNV; gene therapy;
XX KW antifungal; antibacterial; antiviral; plant; enzyme; ds.

XX OS Arabidopsis thaliana.

XX PN WO200206447-A2.

XX PD 24-JAN-2002.

XX PF 18-JUL-2001; 2001WO-US022636.

XX PR 18-JUL-2000; 2000US-0219231P.

XX PA (GEO) GEN HOSPITAL CORP.

XX PI Wildermuth MC, Ausubel FM, Dewdney J;

XX DR WPI; 2002-179788/23.

XX CC Inducing disease resistance on plant to bacterial and fungal pathogens
XX CC e.g. powdery mildew or leaf-spotting bacterial pathogen, comprises
XX CC transforming plant cell with nucleic acid molecule encoding isochorismate
XX CC synthase.

XX PS Disclosure; Page 108-109; 119pp; English.

XX CC The present invention relates to methods of conferring enhanced disease
XX CC resistance on a plant or plant component to a pathogen. The method
XX CC involves introducing a nucleic acid molecule encoding isochorismate
XX CC synthase (ICS) into a plant cell to yield transformed plant cell and
XX CC generating a plant or plant component from transformed plant cell, where
XX CC the nucleic acid molecule is expressed in cells of the plant and confers
XX CC enhanced disease resistance on the plant. Isochorismate synthases are
XX CC useful for synthesizing salicylic acid (SA) which plays an important role
XX CC in systemic acquired resistance (SAR). The methods are used to confer
XX CC enhanced disease resistance on plants or plant components to a pathogen
XX CC including a fungal, especially powdery mildew e.g. Erysiphe orontii or E.
XX CC cichoracearum, or a bacterial pathogen, preferably a leaf-spotting
XX CC bacterial pathogen such as Xanthomonas, Erwinia and a virulent or
XX CC avirulent strain of Pseudomonas syringae. It confers disease resistance
XX CC on a plant to pathogenic nematodes including root-knot nematodes (e.g.
XX CC Meloidogyne sp.), cyst nematodes (e.g. Heterodera sp.), root-attacking
XX CC nematodes (e.g. Rotylenchulus reniformis) and above-ground nematodes
XX CC (e.g. Anguina funesta) and viral pathogens including tobacco mosaic virus
XX CC (TMV), tobacco necrosis virus (TNV), potato leaf roll virus, potato virus
XX CC X, potato virus Y, tomato ring spot virus and tomato spotted wilt virus.
XX CC Sequences of the invention are also used for gene therapy. The present
XX CC sequence is Arabidopsis thaliana isochorismate synthase (At ICS) related
XX CC DNA

XX SQ Sequence 4328 BP; 1456 A; 562 C; 745 G; 1462 T; 0 U; 3 Other;

Query Match 72.3%; Score 18.8; DB 6; Length 4328;
Best Local Similarity 90.9%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 AACACCAATTGATGCTTTAAG 25
|||||
Db 250 AACACCAATTGATGCTTTAAG 229

RESULT 14
AAV15448/c

ID AAV15448 standard; DNA; 4505 BP.

XX AC AAV15448;

XX DT 20-JUL-1998 (first entry)

XX DE Arabidopsis PR-1 promoter.

XX KW Promoter PR-1; salicylic acid, 2,6-dichloroisonicotinic acid;
XX KW benzo(1,2,3)thiadiazole-7-carbothioic acid S-methyl ester;
XX KW transgenic plant; ds.

XX OS Arabidopsis thaliana.

XX PH Key Location/Qualifiers

FT FT promoter 1..4258

FT FT /tag= a

FT FT /note= "full-length PR-1 promoter"

FT FT promoter 2966..4258

FT FT /tag= d

FT FT /note= "PR-1 promoter fragment in pLTD6D, used for LS

FT FT construct construction"

FT FT promoter 3444..4258

FT FT /tag= e

FT FT /note= "PR-1 promoter fragment in pLTD7D, confers similar

FT FT level of chemical induction as full-length promoter"

FT FT primer_bind 3554..3563

FT FT /tag= h

FT FT /note= "LS1"

FT FT promoter 3561..4258

FT FT /tag= f

FT FT /note= "PR-1 promoter fragment in pLTD7D, confers

FT FT chemical induction of gene expression at levels reduced

FT FT by approx. 3-4 fold as compared to full-length promoter"

FT FT primer_bind 3564..3573

FT FT /tag= i

FT FT /note= "LS2"

FT FT primer_bind 3574..3583

FT FT /tag= j

FT FT /note= "LS3"

FT FT primer_bind 3584..3593

FT FT /tag= k

FT FT /note= "LS4"

FT FT primer_bind 3594..3603

FT FT /tag= l

FT FT /note= "LS5"

FT FT primer_bind 3604..3613

FT FT /tag= m

FT FT /note= "LS6"

FT FT primer_bind 3614..3623

FT FT /tag= n

FT FT /note= "LS7"

FT FT primer_bind 3624..3633

FT FT /tag= o

FT FT /note= "LS8"

FT FT primer_bind 3634..3643

FT FT /tag= p

FT FT /note= "LS9"

FT FT promoter 3638..4258

FT FT /tag= g

FT FT /note= "PR-1 promoter fragment in pLTD72D, confers no

```
FT primer_bind      chemical induction of gene expression"
FT 3644. .3653
FT /*tag= q
FT /note= "LS10"
FT primer_bind      3654. .3663
FT /*tag= r
FT /note= "LS11"
FT primer_bind      3664. .3671
FT /*tag= s
FT /note= "LS12"
FT primer_bind      3672. .3681
FT /*tag= t
FT /note= "LS13"
FT TATA_signal      4229. .4232
FT /*tag= b
FT CDS               4294. .4296
FT /*tag= c
FT /note= "start codon for translation"
FT
FT WO9803536-A1.
FT
FT PD 29-JAN-1998.
FT
FT XX 18-JUL-1997; 97WO-US012626.
FT
FT XX 23-JUL-1996; 96US-0027228P.
FT
FT XX (NOVS ) NOVARTIS CORP.
FT
FT XX Lebel EG, Ryals JA, Thorne L, Uknes SJ, Ward ER;
FT WPI; 1998-120690/11.
FT
FT XX New chemically inducible promoter from Arabidopsis - used to regulate
FT gene expression in response to e.g. salicylic acid.
FT
FT PS Claim 1; Page 37-39; 60pp; English.
FT
FT XX This DNA sequence comprises the full-length, chemically-inducible
FT Arabidopsis PR-1 promoter in plasmid pUTDID. Cis-acting regulatory
FT elements in the PR-1 promoter involved in chemical induction have been
FT characterised using deletion and linker-scanning mutagenesis and in vivo
FT footprinting. At least a portion of the region of the promoter between
FT positions -698 and -621 (relative to the transcription start site of the
FT PR-1 gene) is required for induction of gene expression by such chemicals
FT as salicylic acid, 2,6-dichloroisonicotinic acid and
FT benzo(1,2,3)thiadiazole 7-carboxylic acid S-methyl ester. The use of
FT chemically inducible PR-1 promoter fragments to regulate gene expression
FT in plants in the presence of inducing chemicals is disclosed, as well as
FT use of these elements for the isolation of transcriptional regulatory
FT proteins involved in the promoter regulation and for the construction of
FT inducible hybrid promoters. Typical genes which may be regulated are
FT genes that control flowering and fruit ripening, impart tolerance to
FT herbicides or pests, produce enzymes or secondary metabolites, impart
FT sterility or dwarfism, improve nutritional qualities etc., also those
FT CC that express antisense sequences or are involved in production of
FT CC plastics, e.g. polyhydroxybutyrate, or their precursors, perfumes,
FT CC pharmaceuticals etc
FT
FT SQ Sequence 4505 BP; 1501 A; 705 C; 796 G; 1500 T; 0 U; 3 Other;
FT
FT Query Match      72.3%; Score 18.8; DB 2; Length 4505;
FT Best Local Similarity 90.9%; Pred. No. 1.8e+02;
FT Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
FT
FT QY 4 AACACCAATTGATGCTTTAAG 25
FT |||||
FT DB 250 AACACCAATTGATGCTTTAAG 229
FT
FT RESULT 15
FT AAA39651/c
FT ID AAA39651 standard; DNA; 14113 BP.
AC AAA39651;
AC 11-SEP-2000 (first entry)
DE Fusion construct PR-1/luc DNA.
XX PR-1; luciferase; firefly; PR-1/luc; plant disease; resistance; mutant;
KW systemic acquired resistance; SAR; lesion mimic phenotype; antibacterial;
KW antiviral; fungicide; plant breeding program; fungicide screening;
KW pathogen; ss.
XX Arabidopsis thaliana.
OS Photinus pyralis.
OS Synthetic.
XX US6057490-A.
XX 02-MAY-2000.
XX 30-DEC-1998; 98US-00223134.
XX 08-JAN-1993; 93US-00002285.
XX 10-DEC-1993; 93US-00165238.
XX 16-MAY-1996; 96US-00648949.
XX 18-DEC-1997; 97US-0092801.
XX (NOVS ) NOVARTIS FINANCE CORP.
XX Uknes SJ, Ward ER, Maleck K, Ryals JA;
XX WPI; 2000-338627/29.
XX Selecting disease resistant mutant plants useful for plant breeding,
XX involves selecting uninfected transformed plants that constitutively
XX express a reporter gene regulated by SAR gene promoter.
XX Claim 5; Col 25-38; 20pp; English.
XX This invention describes a novel method for selecting disease resistant
XX mutant plants (I) by generating a population of plants transformed with a
XX reporter gene (RG) under the regulation of systemic acquired resistance
XX (SAR) gene promoter and selecting uninfected phenotypically normal plants
XX that are lacking a lesion mimic phenotype and constitutively expressing
XX RG, in the absence of viral, bacterial or fungal infections. The products
XX of the invention have antibacterial, antiviral and fungicidal activity.
XX Selected disease resistant mutant plants are useful in plant breeding
XX programs. They also have utility in disease and pathogenesis testing and
XX fungicide screening programs. This sequence represents a fusion construct
XX PR-1/luc DNA which is constructed from Arabidopsis thaliana PR-1 gene and
XX the firefly (Photinus pyralis) luciferase gene, and is used in the method
XX of the invention
XX SQ Sequence 14113 BP; 3645 A; 3225 C; 3345 G; 3766 T; 0 U; 132 Other;
XX
XX Query Match      72.3%; Score 18.8; DB 3; Length 14113;
XX Best Local Similarity 90.9%; Pred. No. 2.2e+02;
XX Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 4 AACACCAATTGATGCTTTAAG 25
XX |||||
XX DB 7972 AACACCAATTGATGCTTTAAG 7951
XX
XX Search completed: November 26, 2005, 13:40:52
XX Job time : 18.6769 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November, 26, 2005, 12:58:53 ; Search time 145.336 Seconds
(without alignments)
8370.011 Million cell updates/sec

Title: US-09-555-529-19

Perfect score: 26

Sequence: 1 gcgaacccaatttgatgctttaaga 26

Scoring table: IDENTITY_NUC

Gapop 10,0 , Gapext 1.0

Searched: 41078325, seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: gb_est1: *
2: gb_est2: *
3: gb_est3: *
4: gb_hic: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_est7: *
9: gb_gss1: *
10: gb_gss2: *
11: gb_gss3: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	26	100.0	249	1	AI493544 th36a10.x
2	26	100.0	328	2	BE091046 PM4-BT072
3	26	100.0	384	1	AA770446 ah89e07.8
4	26	100.0	384	1	AI493563 th36c10.x
5	26	100.0	386	3	BM661987 UI-E-CK1
6	26	100.0	469	1	AI087818 o624a09.x
7	26	100.0	476	1	AI0878396 tc78f05.x
8	26	100.0	485	1	AI089251 qb09a11.x
9	26	100.0	488	3	BM689918 UI-E-CK1
10	26	100.0	512	5	BQ548883 ik93a03.y
11	26	100.0	516	2	BG944189 ax47e12.y
12	26	100.0	521	2	BE349628 ht58g10.x
13	26	100.0	543	6	CB158644 K-EST0218
14	26	100.0	575	3	BM750112 K-EST0025
15	26	100.0	591	2	B8896845 601437548
16	26	100.0	592	8	CX866354 HESCA.10
17	26	100.0	605	6	CB242699 UI-CF-FN0
18	26	100.0	660	6	AL597250 DFE2p313A
19	26	100.0	682	1	AV721396 AV721396
20	26	100.0	698	1	AI650375 wa90b01.x
21	26	100.0	701	1	AI807250 wf38f09.x
22	26	100.0	720	7	CK301009 UI-E-EJ1-

23	26	100.0	733	5	BX104466
24	26	100.0	848	1	AL558810
25	26	100.0	1339	4	CR618602
26	26	100.0	1396	4	CR595908
27	25	96.2	723	3	BQ186497
28	24.4	93.8	259	6	CF526711
29	24.4	93.8	349	1	AA805923
30	24.4	93.8	558	5	BU737321
31	24.4	93.8	679	1	AM028238
32	24.4	93.8	691	1	AM028686
33	24.4	93.8	703	1	AM035210
34	24.4	93.8	776	2	BF571691
35	24.4	93.8	786	1	AJ819816
36	24.4	93.8	864	1	AM037867
37	24.4	93.8	871	1	AJ819723
38	24.4	93.8	949	3	BM453332
39	23	88.5	477	1	AJ556645
40	23	88.5	624	5	BQ602324
41	23	88.5	663	1	AJ660240
42	23	88.5	682	6	CF795929
43	23	88.5	1543	4	AY609991
44	22.8	87.7	800	1	AM038283
45	22	84.6	398	8	H75516

ALIGNMENTS

RESULT 1
AI493544
LOCUS th36a10.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2120346 3', linear EST 30-MAR-1999
DEFINITION mRNA sequence.
ACCESSION AI493544
VERSION AI493544.1 GI:4394547
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 249)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapsb@mail.nih.gov
Life Technologies catalog #: 11548-013
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 841 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 246.
Location/Qualifiers
1. 249
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2120346"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Pan1"
/notes="Organ: pancreas; Vector: pCMV-SPORT6; Site: 1; Salt: Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.72 kb. Life Technologies catalog #: 11548-013"

FEATURES
source
1. 249
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2120346"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Pan1"
/notes="Organ: pancreas; Vector: pCMV-SPORT6; Site: 1; Salt: Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.72 kb. Life Technologies catalog #: 11548-013"

ORIGIN
Query Match 100.0%; Score 26; DB 1; Length 249;
Best Local Similarity 100.0%; Pred. No. 0.42;

Life Technologies catalog #: 11548-013
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 843 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 366.
 Location/Qualifiers

FEATURES

source
 1. .384
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2120370"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B"
 /clone_lib="NCI CGAP Panl"
 /note="Organ: pancreas; Vector: pCMV-SPORT6; Site 1: SalI;
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.72 kb. Life Technologies catalog #:
 11548-013"

ORIGIN

Query Match 100.0%; Score 26; DB 1; Length 384;
 Best Local Similarity 100.0%; Pred. No. 0.45;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGACACCAATTTGATGCTTTAAGA 26
 |||||
 Db 221 GCGACACCAATTTGATGCTTTAAGA 246

RESULT 5

BM661987
 LOCUS
 DEFINITION UI-E-CK1-abo-f-09-0-UI.s1 UI-E-CK1 Homo sapiens cDNA clone
 UI-E-CK1-abo-f-09-0-UI 3', mRNA sequence.

ACCESSION BM661987.1 GI:18966162
 VERSION
 KEYWORDS
 SOURCE EST.

ORGANISM

Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homnidae; Homo.

REFERENCE 1 (bases 1 to 386)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

PUBMED 8889548
 COMMENT Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu

Tissue Procurement: Dr. Gregg Hageman
 CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 Seq primer: M13 Forward

POLYA=Yes.
 Location/Qualifiers

FEATURES

source
 1. .386
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-E-CK1-abo-f-09-0-UI"
 /tissue_type="Retina Foveal and Macular"

/dev stage="adult"
 /lab host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_lib="UI-E-CK1"
 /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I;
 UI-E-CK1 is a normalized cDNA library containing the
 following tissue(s): Retina Foveal and Macular. The
 library was constructed according to Bonaldo, Lennon and
 Soares, Genome Research, 6:791-806, 1996. First strand
 cDNA synthesis was primed with an oligo-dT primer
 containing a Not I site. Double stranded cDNA was ligated
 to an EcoR I adaptor, digested with Not I, and cloned
 directionally into pT7T3-Pac vector. The oligonucleotide
 used to prime the synthesis of first-strand cDNA contains
 a library tag sequence that is located between the Not I
 site and the (dT)18 tail. The sequence tag for this
 library is GTCC. This library was created for the program,
 Gene Discovery in the Visual System, supported by National
 Eye Institute (NEI).
 TAG TISSUE=Foveal and Macular Retina
 TAG_LIB=UI-E-CK1
 TAG_SEQ=GTCC"

ORIGIN

Query Match 100.0%; Score 26; DB 3; Length 386;
 Best Local Similarity 100.0%; Pred. No. 0.45;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGACACCAATTTGATGCTTTAAGA 26
 |||||
 Db 238 GCGACACCAATTTGATGCTTTAAGA 263

RESULT 6

AI087818
 LOCUS
 DEFINITION Oo24a09.x1 Soares NSF F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
 IMAGE:1567096 3', mRNA sequence.

ACCESSION AI087818
 VERSION
 KEYWORDS
 SOURCE EST.

ORGANISM

Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homnidae; Homo.

REFERENCE 1 (bases 1 to 469)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL

COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapsb@mail.nih.gov
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 846 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 416.

Location/Qualifiers

FEATURES

source

1. .469
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1567096"
 /lab host="DH10B"
 /clone_lib="Soares NSF F8_9W_OT_PA_P_S1"
 /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
 a modified polylinker; Site 1: Not I; Site 2: Eco RI;
 Equal amounts of plasmid DNA from five normalized
 libraries were mixed, and es circles were made in vitro.
 Following HAP purification, this DNA was used as tracer in
 a subtractive hybridization reaction. The driver was
 PCR-amplified cDNAs from pools of 5,000 clones made from

the same 5 libraries. The pools consisted of the following libraries and clones: Soares NBHP pool 1: 309384-310919, 323208-325895 Soares NB2HP pool 1: 145032-147335, 147720-148103, 148872-149255, 15002 - 150407, 151176-152327 Soares NB2HP8-9W pool 1: 758280-760583, 772104-774407 Soares NBHPA pool 1: 304776-306311, 320136-322823, 326280-326663 Soares NBHOT pool 1: 723720-726407, 739080-740999 Subtraction by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 100.0%; Score 26; DB 1; Length 469;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGAACCAATTTGATGCTTTAAGA 26

Db 231 GCGAACCAATTTGATGCTTTAAGA 256

RESULT 7

AI378396

LOCUS tc78f05.xl Soares NBHPu_S1 Homo sapiens cDNA clone IMAGE:2070753
DEFINITION 3', mRNA sequence.

ACCESSION AI378396

VERSION AI378396.1 GI:4188249

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

REFERENCE 1 (bases 1 to 476)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 625 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 442.

FEATURES

Location/Qualifiers

1..476

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:2070753"

/tissue_type="Pooled human melanocyte, fetal heart, and

pregnant uterus"

/lab_host="DH10B"

/clone_lib="Soares NBHPu S1"

/note="Organ: mixed (see below); Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NBHM, pregnant uterus NBHPu, and fetal heart NBHH19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."

ORIGIN

Query Match 100.0%; Score 26; DB 1; Length 476;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGAACCAATTTGATGCTTTAAGA 26

Db 218 GCGAACCAATTTGATGCTTTAAGA 243

RESULT 8

AI089251

LOCUS qB09a1.xl Soares_pregnant_uterus_NbHPu Homo sapiens cDNA clone
DEFINITION IMAGE:1695740 3', mRNA sequence.

ACCESSION AI089251

VERSION AI089251.1 GI:3428310

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

REFERENCE 1 (bases 1 to 485)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 442.

FEATURES

Location/Qualifiers

1..485

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:1695740"

/sex="female"

/dev_stage="adult"

/lab_host="DH10B"

/clone_lib="Soares_pregnant_uterus_NbHPu"

/note="Organ: uterus; Vector: pT7T3-Pac; Site 1: Not I;

Site 2: Eco RI; 1st strand cDNA was primed with a Not I -

oligo(dT) primer [5'

AAC TGAAGATTCGCGCGCTTTT TTTT TTTT 3'],

double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT7T3 vector. Library

went through one round of normalization. Library

constructed by M. Fatima Bonaldo."

ORIGIN

Query Match 100.0%; Score 26; DB 1; Length 485;

Best Local Similarity 100.0%; Pred. No. 0.47;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGAACCAATTTGATGCTTTAAGA 26

Db 218 GCGAACCAATTTGATGCTTTAAGA 243

RESULT 9

BM689918/c

LOCUS UI-E-CKI-abo-f-09-0-UI.r1 UI-E-CKI Homo sapiens cDNA clone
DEFINITION UI-E-CKI-abo-f-09-0-UI 5', mRNA sequence.

ACCESSION BM689918

VERSION BM689918.1 GI:19003176

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

REFERENCE 1 (bases 1 to 488)

AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

PUBLISHED 8889548

COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRP, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
Seq primer: M13 Reverse.

FEATURES Location/Qualifiers
1..488
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-CK1-abo-f-09-0-UI"
/tissue_type="Retina Foveal and Macular"
/dev_stage="adult"
/lab_host="DHI0B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-CK1"
/notes="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-CK1 is a normalized cDNA library containing the following tissue(s): Retina Foveal and Macular. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GRCC. This library was created for the program, Gene discovery in the Visual System, supported by National Eye Institute (NEI)."

ORIGIN
Query Match 100.0%; Score 26; DB 3; Length 488;
Best Local Similarity 100.0%; Pred. No. 0.47; Mismatches 0; Indels 0; Gaps 0;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCGAACCAATTTGATGCTTTAAGA 26
|||||
Db 463 GCGAACCAATTTGATGCTTTAAGA 438

RESULT 10
BQ548883/c
LOCUS BQ548883
DEFINITION ik93a03.y1 Human insulinoma Homo sapiens cDNA clone IMAGE:6027917 5', similar to, TR:O60870 O60870 KIN17 PROTEIN. ;, mRNA sequence.
ACCESSION BQ548883
VERSION BQ548883.1 Gi:21433386
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 512)
Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K., Lemishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Narra, M., Pape, D., Wylie, T., Martin, J., Blistain, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J.,

TITLE Cardenas, M., Gibbons, M., McCann, R., Cole, R., Teagareishvili, R., Williams, T., Jackson, Y. and Bowers, Y.

JOURNAL Endocrine Pancreas Consortium

COMMENT Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohpc.harvard.edu
Library was constructed by Dr. J. Ferrer In vivo mass-excised to pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Seq primer: -40RP from Gibco
High quality sequence stop: 434.

FEATURES Location/Qualifiers
1..512
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6027917"
/tissue_type="insulinoma"
/lab_host="DHI0B (phage-resistant)"
/clone_lib="Human insulinoma"
/notes="Organ: pancreas; Vector: pBluescript SK-; Site 1: XhoI; Site 2: EcoRI; Constructed with lambda ZAPII system (Stratagene) by Dr. J. Ferrer, in vivo mass-excised to pBluescript SK- by Dr. H. Inoue following the Washington University protocol (http://genome.wustl.edu/est/lambda_protocol.shtml). Please contact Hiroshi Inoue, MD/PhD for further information on this library (Metabolism Division, Permutt Laboratory, Washington University School of Medicine, Box 8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this is a Washington University Pancreas EST project library."

ORIGIN
Query Match 100.0%; Score 26; DB 5; Length 512;
Best Local Similarity 100.0%; Pred. No. 0.47; Mismatches 0; Indels 0; Gaps 0;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCGAACCAATTTGATGCTTTAAGA 26
|||||
Db 360 GCGAACCAATTTGATGCTTTAAGA 335

RESULT 11
BQ944189
LOCUS BQ944189
DEFINITION ax47e12.x1 Hembase; Erythroid Progenitor Cells (LCB:ax library) Homo sapiens cDNA clone ax47e12 random, mRNA sequence.
ACCESSION BQ944189
VERSION BQ944189.1 GI:14343561
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 516)
Gubin, A.N., Njoroge, J.M., Bouffard, G.G. and Miller, J.L.
Gene expression in proliferating human erythroid cells
Genomics 59 (2), 168-177 (1999)
10409428
Contact: Jeffery L. Miller
Laboratory of Chemical Biology
National Institute of Diabetes and Digestive and Kidney Diseases
Building 10, Room 9B17, National Institutes of Health, Bethesda, MD 20892, USA
Tel: 301 402 2373

Fax: 301 435 5148
 Email: jmf@nih.gov
 DNA Sequencing and analyses by National Institutes of Health
 Intramural Sequencing Center (NISC).
 Plate: 47 row: e column: 12
 Seq primer: -21M13 forward primer (ABI).
 Location/Qualifiers

FEATURES

1..516
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="ax47e12"
 /sex="unknown"
 /tissue_type="blood"
 /cell_type="Erythroid cells"
 /cell_line="Primary Culture of Peripheral Blood
 Mononuclear Cells"
 /dev_stage="progenitor; EPO responsive CD71++++"
 /lab_host="SOLR"
 /clone_lib="Hembase; Erythroid Progenitor Cells (LCB:ax
 library)"
 /note="Organ: blood; Vector: Lambda ZAP II; Site 1: EcoRI;
 site 2: EcoRI; 65,000 proliferating erythroid cells from
 the Buffy coat of a blood donation were obtained by flow
 cytometric separation after a 5-day culture period in the
 presence of erythropoietin. Total RNA was purified from
 the sorted cell population using Trizol reagent. RNA (0.3
 ug) was converted into double stranded cDNA using
 Clontech's CapFinder cDNA Library Construction Kit
 (Clontech) according to the manufacturer's protocol and
 cloned into EcoRI digested Lambda Zap II vector
 (Stratagene). The phage library was amplified once prior
 to in vivo excision in SOLR cells. Individual colonies
 were grown, and the cDNA inserts were sequenced in high
 throughput (NIH intramural sequencing center
 http://www.nisc.nih.gov/)."

Query Match 100.0%; Score 26; DB 2; Length 516;
 Best Local Similarity 100.0%; Pred. No. 0.47;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGAACACCAATTGATGCTTTAAGA 26
 |||||
 DB 255 GCGAACACCAATTGATGCTTTAAGA 280

ORIGIN

Query Match 100.0%; Score 26; DB 2; Length 516;
 Best Local Similarity 100.0%; Pred. No. 0.47;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGAACACCAATTGATGCTTTAAGA 26
 |||||
 DB 255 GCGAACACCAATTGATGCTTTAAGA 280

RESULT 12

BE349628
 LOCUS
 DEFINITION
 ht58g10.x1 NCI CGAP Mell15 Homo sapiens cDNA clone IMAGE:3150978 3',
 similar to TR:060870 O60870 KIN17 PROTEIN. ;, mRNA sequence.

ACCESSION BE349628

VERSION BE349628.1 GI:9261481

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Hominidae; Homo.

1 (bases 1 to 521)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgaps-r@mail.nih.gov

Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life

Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The

I.M.A.G.E. Consortium DNA Sequencing by: Washington University

Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL, send email to:
 info@image.llnl.gov
 Seq primer: -40UP from Gibco
 High quality sequence stop: 439.
 Location/Qualifiers

FEATURES

1..521
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3150978"
 /tissue_type="malignant melanoma, metastatic to lymph
 node"
 /lab_host="DH10B"
 /clone_lib="NCI CGAP Mell15"
 /note="Organ: skin; Vector: pCMV-SPORT6; Site 1: SalI;
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies."

ORIGIN

Query Match 100.0%; Score 26; DB 2; Length 521;
 Best Local Similarity 100.0%; Pred. No. 0.47;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGAACACCAATTGATGCTTTAAGA 26
 |||||
 DB 229 GCGAACACCAATTGATGCTTTAAGA 254

RESULT 13

CB158644/c

LOCUS

DEFINITION

X-EST0218017 L18POOLn1 Homo sapiens cDNA clone L18POOLn1-12-F09

5', mRNA sequence.

ACCESSION CB158644

VERSION CB158644.1 GI:28143780

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Hominidae; Homo.

1 (bases 1 to 543)

Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,

Kim,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and

Kim,Y.S.

21C Frontier Korean EST Project 2001

Unpublished (2002)

Contact: Kim YS

Genome Research Center

Korea Research Institute of Bioscience & Biotechnology

52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea

Tel: +82-42-860-4470

Fax: +82-42-860-4409

Email: yongseung@mail.kribb.re.kr

Plate: 12 row: F column: 09

High quality sequence stop: 543.

FEATURES

source

1..543

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="L18POOLn1-12-F09"

/cell_line="SNU-354+Cho-CK+Choi-CK+HLK-3"

/lab_host="Top10F"

/clone_lib="L18POOLn1"

/note="Organ: Liver; Vector: pT73-Pac; Site 1: EcoRI;

Site 2: NotI; The library was contributed by the Soares

laboratory and it was constructed as described by Bonaldo,

M.F., Lennon, G. and Soares, M.B. (1996), Genome Research

6(9): 791-806. RNA was prepared from harvested cell

culture."

ORIGIN

Query Match 100.0%; Score 26; DB 6; Length 543;
 Best Local Similarity 100.0%; Pred. No. 0.47; Indels 0; Gaps 0;
 Matches 26; Conservative 0; Mismatches 0;

Qy 1 GCGAACACCAATTGATGCTTTAAGA 26
 |||||
 Db 468 GCGAACACCAATTGATGCTTTAAGA 443

RESULT 14
 BM750112/c
 LOCUS 575 bp mRNA linear EST 04-MAR-2002
 DEFINITION K-EST0025570 S5SNU484 Homo sapiens cDNA clone S5SNU484-4-G04 5',
 mRNA sequence.

ACCESSION BM750112
 VERSION BM750112.1 GI:19079730
 KEYWORDS EST.
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homiidae; Homo.

REFERENCE 1 (bases 1 to 575)
 Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
 Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and
 Kim, Y.S.

TITLE 21C Frontier Korean EST Project 2001
 JOURNAL Unpublished (2002)
 COMMENT Contact: Kim YS

Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409

Email: yongseung@mail.kribb.re.kr
 Plate: 4 row: G column: 04
 High quality sequence stop: 575.

FEATURES
 Location/Qualifiers

1..575
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S5SNU484-4-G04"
 /sex="M"
 /tissue_type="Stomach"
 /cell_line="Epithelial"
 /cell_line="SNU-484"
 /lab_host="Top10P"
 /clone_lib="S5SNU484"
 /note="Organ: Stomach; Vector: pTZ19RP1; Site 1: EcoRI;
 Site 2: NotI; The poly (A)+ RNA was decapped with tobacco
 acid pyrophosphatase (TAP) and ligated with DNA-RNA linker
 including EcoRI site by treatment of T4 RNA ligase. The
 first strand cDNA was synthesized from oligo dt-selected
 mRNA by priming with dt-tailed vector. The dt-tailed
 vector was adjusted to have about 60nt. The cDNA vector
 was circularized with E. coli DNA ligase after digestion
 of EcoRI which site is also included in vector. An RNA
 strand converted to a DNA strand by Okayama-Berg method.
 The obtained cDNA vectors were used for transformation of
 competent cells E. coli Top10P by electroporation
 method."

ORIGIN

Query Match 100.0%; Score 26; DB 3; Length 575;
 Best Local Similarity 100.0%; Pred. No. 0.48; Indels 0; Gaps 0;
 Matches 26; Conservative 0; Mismatches 0;

Qy 1 GCGAACACCAATTGATGCTTTAAGA 26
 |||||
 Db 371 GCGAACACCAATTGATGCTTTAAGA 346

RESULT 15
 BE896845/c
 LOCUS 591 bp mRNA linear EST 20-OCT-2000
 DEFINITION 601437548F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3922787 5',
 mRNA sequence.

ACCESSION BE896845
 VERSION BE896845.1 GI:10361687
 KEYWORDS EST.
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homiidae; Homo.

REFERENCE 1 (bases 1 to 591)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 JOURNAL
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs@mail.nih.gov

Tissue Procurement: ATCC/DCTD/BTP
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov

Plate: LLAM9757 row: 0 column: 12
 High quality sequence stop: 587.

FEATURES
 Location/Qualifiers

1..591
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3922787"
 /tissue_type="melanotic melanoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 72"
 /note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 2 kb. Library constructed by Life
 Technologies."

ORIGIN

Query Match 100.0%; Score 26; DB 2; Length 591;
 Best Local Similarity 100.0%; Pred. No. 0.48; Indels 0; Gaps 0;
 Matches 26; Conservative 0; Mismatches 0;

Qy 1 GCGAACACCAATTGATGCTTTAAGA 26
 |||||
 Db 588 GCGAACACCAATTGATGCTTTAAGA 563

Search completed: November 27, 2005, 00:57:55
 Job time : 147.336 secs

This Page Blank (uspto)


```
;
; APPLICANT: Maleck, Klaus
; TITLE OF INVENTION: Method For Breeding Disease Resistance
; TITLE OF INVENTION: Into Plants
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 605749artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/223,134
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/992,801
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/165,248
; FILING DATE: 10-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/002,285
; FILING DATE: 08-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC1673/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14113 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "PR-1/luc construct"
;
US-09-223-134-1
```

```
Query Match 72.3%; Score 18.8; DB 3; Length 14113;
Best Local Similarity 90.9%; Pred. No. 56;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 4 AACACCAATTGATGCTTTAAG 25
|||||
Db 7972 AACACCAATTGATGCTTTAAG 7951
```

```
RESULT 3
US-09-992-801-1/c
; Sequence 1, Application US/08992801
; Patent No. 6107544
; GENERAL INFORMATION:
; APPLICANT: Ryals, John
; APPLICANT: Uknes, Scott
; APPLICANT: Ward, Eric
; APPLICANT: Delaney, Terry
; APPLICANT: Lawton, Kay
; APPLICANT: Weymann, Kris
; APPLICANT: Steiner, Henry-York
; APPLICANT: Maleck, Klaus
; TITLE OF INVENTION: Method For Breeding Disease Resistance
; TITLE OF INVENTION: Into Plants
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6107544artis Corporation
```

```
;
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/992,801
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/648,949
; FILING DATE: 16-MAY-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/165,248
; FILING DATE: 10-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/002,285
; FILING DATE: 08-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC1673/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14113 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "PR-1/luc construct"
;
US-08-992-801-1

Query Match 72.3%; Score 18.8; DB 3; Length 14113;
Best Local Similarity 90.9%; Pred. No. 56;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 AACACCAATTGATGCTTTAAG 25
|||||
Db 7972 AACACCAATTGATGCTTTAAG 7951

RESULT 4
US-09-223-535-1/c
; Sequence 1, Application US/09223535
; Patent No. 6232525
; GENERAL INFORMATION:
; APPLICANT: Ryals, John
; APPLICANT: Uknes, Scott
; APPLICANT: Ward, Eric
; APPLICANT: Delaney, Terry
; APPLICANT: Lawton, Kay
; APPLICANT: Weymann, Kris
; APPLICANT: Steiner, Henry-York
; APPLICANT: Maleck, Klaus
; TITLE OF INVENTION: Method For Breeding Disease Resistance
; TITLE OF INVENTION: Into Plants
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6232525artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/223,535
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/992,801
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/165,248
FILING DATE: 10-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/002,285
FILING DATE: 08-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC1673/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1413 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "PR-1/luc construct"
US-09-223-535-1

Query Match 72.3%; Score 18.8; DB 3; Length 1413;
Best Local Similarity 90.9%; Pred. No. 56;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 AACACCAATTGGATGCTTTAAG 25
|||||
Db 7972 AACACCAATTGGATGCTTTAAG 7951

RESULT 5
US-09-621-976-10260/c
Sequence 10260, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 10260
LENGTH: 381
TYPE: DNA
ORGANISM: Homo sapiens
US-09-621-976-10260

Query Match 70.8%; Score 18.4; DB 3; Length 381;
Best Local Similarity 95.0%; Pred. No. 48;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 AACACCAATTGGATGCTTTA 23
|||||
Db 119 AACACCAATTGGATGCTTTA 100

RESULT 6

US-09-513-999C-34547
Sequence 34547, Application US/09513999C
Patent No. 6783961
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Duclert, A.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
Patent No. 6783961
FILE REFERENCE: 59 US2,REG
CURRENT APPLICATION NUMBER: US/09/513,999C
CURRENT FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/122,487
PRIOR FILING DATE: 1999-02-26
NUMBER OF SEQ ID NOS: 36681
SOFTWARE: Patent.pm
SEQ ID NO 34547
LENGTH: 217
TYPE: DNA
ORGANISM: Homo sapiens
US-09-513-999C-34547

Query Match 69.2%; Score 18; DB 3; Length 217;
Best Local Similarity 80.8%; Pred. No. 67;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GCGAACACCAATTGGATGCTTTAAGA 26
|||||
Db 82 GCGAACACCAATTGGATGCTTTAAGA 107

RESULT 7
US-08-651-155B-42
Sequence 42, Application US/08651155B
Patent No. 6365401
GENERAL INFORMATION:
APPLICANT: Mahan Dr., Michael J.
APPLICANT: Conner Mr., Christopher P.
APPLICANT: Hiethoff Mr., Douglas M.
TITLE OF INVENTION: METHOD AND PROBES FOR THE IDENTIFICATION
TITLE OF INVENTION: OF MICROBIAL GENES SPECIFICALLY INDUCED DURING HOST
NUMBER OF SEQUENCES: 255
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chrismen, Bynum & Johnson, P.C.
STREET: 1900 Fifteenth Street
CITY: Boulder
STATE: CO
COUNTRY: USA
ZIP: 80302
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/651,155B
FILING DATE: 17-MAY-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Petersen Mr., Steven C.
REGISTRATION NUMBER: 36,238
REFERENCE/DOCKET NUMBER: 17060.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/546-1300
TELEFAX: 303/449-5426
TELEX: ABA1475
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 302 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

; Sequence 1495, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543.681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 1495
; LENGTH: 672
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-1495

Query Match 67.7%; Score 17.6; DB 3; Length 672;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GCGAACCAATTTGATGCTTTAA 24
||||| ||||| ||||| ||||| |||||
Db 224 GCGAACAGCAAGTTGATGTTAA 247
||||| ||||| ||||| ||||| |||||

RESULT 12
US-08-956-171E-659
; Sequence 659, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956.171E
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 659:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1899 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double

; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 659:
US-08-956-171E-659

Query Match 67.7%; Score 17.6; DB 3; Length 1899;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 GAACACCAATTTGATGCTTTAAGA 26
||||| ||||| ||||| ||||| |||||
Db 1302 GAACACCAATTTGATGCTTTAAGA 1325
||||| ||||| ||||| ||||| |||||

RESULT 13
US-08-781-986A-659
; Sequence 659, Application US/08781986A
; Patent No. 6737248
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 659:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1899 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-781-986A-659

Query Match 67.7%; Score 17.6; DB 3; Length 1899;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 GAACACCAATTTGATGCTTTAAGA 26
||||| ||||| ||||| ||||| |||||
Db 1302 GAACACCAATTTGATGCTTTAAGA 1325
||||| ||||| ||||| ||||| |||||

RESULT 14
US-09-248-796A-1868
; Sequence 1868, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132

```
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 1868
; LENGTH: 2094
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-1868

Query Match      67.7%; Score 17.6; DB 3; Length 2094;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      3 GAACACCAATTGATGCTTTAAGA 26
Db      1897 GAACAACAACTAGAGGCTTTAAGA 1920

RESULT 15
US-09-949-016-14647/c
; Sequence 14647, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14647
; LENGTH: 15125
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(15125)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14647

Query Match      67.7%; Score 17.6; DB 3; Length 15125;
Best Local Similarity 83.3%; Pred. No. 2.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      3 GAACACCAATTGATGCTTTAAGA 26
Db      14672 GAACACAGATTGATGTTTAATA 14649

Search completed: November 27, 2005, 01:10:49
Job time : 7.38437 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:00:41 ; Search time 29.9737 Seconds
(without alignments)
7173.088 Million cell updates/sec

Title: US-09-555-529-19

Perfect score: 26

Sequence: 1 ggaacacacaaattgatgctttaaga 26

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA Main:

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	26	100.0	591	5	US-10-106-698-1187
C 2	19.8	76.2	556	7	US-10-424-599-49273
C 3	19.8	76.2	2002	8	US-10-425-115-18276
4	19.6	75.4	915	5	US-10-012-819-67
5	19.6	75.4	954	7	US-10-335-977-1576
6	19.6	75.4	999	7	US-10-335-977-1577
7	19.6	75.4	1116	3	US-09-882-227-439
C 8	19.2	73.8	3186778	5	US-10-027-632-174961
C 9	19.2	73.8	3186778	6	US-10-027-632-174961
C 10	18.8	72.3	578	4	US-09-925-065A-593526
C 11	18.8	72.3	603	4	US-09-925-065A-722502
C 12	18.8	72.3	4328	3	US-09-908-299-4
C 13	18.4	70.8	656	4	US-09-925-065A-679116
C 14	18.2	70.0	548	4	US-09-925-065A-527754
C 15	18.2	70.0	548	4	US-09-925-065A-527755
C 16	18.2	70.0	548	4	US-09-925-065A-527756
17	18.2	70.0	611	7	US-10-424-599-141373
18	18.2	70.0	2253	6	US-10-171-404A-1
19	18	69.2	65	10	US-11-087-804-11
C 20	18	69.2	65	10	US-11-087-804-13
C 21	18	69.2	158	8	US-10-425-115-39464
22	18	69.2	516	7	US-10-021-323-11567
23	18	69.2	528	7	US-10-021-323-16963

C 24	69.2	559	7	US-10-021-323-5883	Sequence 5883, Ap
C 25	69.2	559	7	US-10-021-323-15438	Sequence 15438, A
C 26	69.2	560	4	US-09-925-065A-902359	Sequence 902359, A
C 27	69.2	574	7	US-10-021-323-11480	Sequence 11480, A
C 28	69.2	599	7	US-10-021-323-17390	Sequence 17390, A
C 29	69.2	606	7	US-10-021-323-17306	Sequence 17306, A
C 30	69.2	1839	9	US-10-367-057-194	Sequence 194, App
C 31	69.2	4498	7	US-10-425-114-34922	Sequence 34922, A
C 32	69.2	5134	8	US-10-425-115-106105	Sequence 106105, A
C 33	69.2	15811	8	US-10-723-860-2720	Sequence 2720, Ap
C 34	69.2	721377	9	US-10-461-862-163	Sequence 163, App
C 35	69.2	583	7	US-10-424-599-21922	Sequence 21922, A
C 36	69.2	419	4	US-09-925-065A-232560	Sequence 232560, A
C 37	69.2	460	4	US-09-925-065A-129311	Sequence 129311, A
C 38	69.2	509	4	US-09-925-065A-737551	Sequence 737551, A
C 39	69.2	530	4	US-09-925-065A-382903	Sequence 382903, A
C 40	69.2	530	4	US-09-925-065A-382904	Sequence 382904, A
C 41	69.2	530	4	US-09-925-065A-382905	Sequence 382905, A
C 42	69.2	530	4	US-09-925-065A-382906	Sequence 382906, A
C 43	69.2	566	4	US-09-925-065A-181407	Sequence 181407, A
C 44	69.2	584	4	US-09-925-065A-21421	Sequence 21421, A
C 45	69.2	600	9	US-10-972-079-85617	Sequence 85617, A

ALIGNMENTS

RESULT 1

US-10-106-698-1187/c
; Sequence 1187, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 1187
; LENGTH: 591
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (587)..(587)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-1187

Query Match 100.0%; Score 26; DB 5; Length 591;
Best Local Similarity 100.0%; Pred. No. 0.25; Indels 0; Gaps 0;
Matches 26; Conservative 0; Mismatches 0

Qy 1 GCGAACCAACCAATTTGATGCTTTAAGA 26

Db 326 GCGAACCAACCAATTTGATGCTTTAAGA 301

RESULT 2

US-10-424-599-49273/c
; Sequence 49273, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei

```
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 49273
; LENGTH: 556
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_15500C.1
US-10-424-599-49273

Query Match          76.2%; Score 19.8; DB 7; Length 556;
Best Local Similarity 91.3%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGACACCAATTGATGCTTTA 23
Db 172 GCGTACCAATTGATGCTTCA 150

RESULT 3
US-10-425-115-18276
; Sequence 18276, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 18276
; LENGTH: 2002
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_116668C.1
US-10-425-115-18276

Query Match          76.2%; Score 19.8; DB 8; Length 2002;
Best Local Similarity 91.3%; Pred. No. 1.9e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 AACACCAATTGATGCTTTAAGA 26
Db 175 AACACCAATTGATGCTTCTGA 197

RESULT 4
US-10-012-819-67
; Sequence 67, Application US/10012819
; Publication No. US20030017478A1
; GENERAL INFORMATION:
; APPLICANT: Legrain, Pierre
; APPLICANT: Selig, Luc
; APPLICANT: Rain, Jean-Christophe
; TITLE OF INVENTION: Collection of Prokaryotic DNA for Two-Hybrid System, Helicobacter
; FILE OF INVENTION: pylori Protein-Protein Interactions and Applications thereof
; FILE REFERENCE: B5053
; CURRENT APPLICATION NUMBER: US/10/012,819
; CURRENT FILING DATE: 2001-10-30
; PRIOR APPLICATION NUMBER: EP 99401066.8
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 278
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 67
```

```
; LENGTH: 915
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(915)
US-10-012-819-67

Query Match          75.4%; Score 19.6; DB 5; Length 915;
Best Local Similarity 84.6%; Pred. No. 2e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCGAACCAATTGATGCTTTAAGA 26
Db 536 GCGATCTCAAAATGATGCTTTAAGA 561

RESULT 5
US-10-335-977-1576
; Sequence 1576, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 1576:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 954 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3) LOCATION 1...954
; SEQUENCE DESCRIPTION: SEQ ID NO: 1576:
US-10-335-977-1576

Query Match          75.4%; Score 19.6; DB 7; Length 954;
Best Local Similarity 84.6%; Pred. No. 2e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

Qy 1 GCGAACACCAATTTGATGCTTTAAGA 26
|||||
Db 686 GCGATCTCAAAATTTGATGCTTTAAGA 711

RESULT 6
US-10-335-977-1577
; Sequence 1577, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 1577:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 998 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (8) LOCATION 1...999
; SEQUENCE DESCRIPTION: SEQ ID NO: 1577:
US-10-335-977-1577
Query Match 75.4%; Score 19.6; DB 7; Length 999;
Best Local Similarity 84.6%; Pred. No. 2e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GCGAACACCAATTTGATGCTTTAAGA 26
|||||
Db 617 GCGATCTCAAAATTTGATGCTTTAAGA 642

RESULT 7
US-09-882-227-439
; Sequence 439, Application US/09882227
; Publication No. US20030158396A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal

; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Coomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20030158396A1a1 Helicobacter Polypeptides in the
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/047002
; CURRENT APPLICATION NUMBER: US/09/882,227
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/902,615
; PRIOR FILING DATE: 1997-07-29
; NUMBER OF SEQ ID NOS: 638
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 439
; LENGTH: 1116
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (70) ... (1065)
US-09-882-227-439
Query Match 75.4%; Score 19.6; DB 3; Length 1116;
Best Local Similarity 84.6%; Pred. No. 2.1e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GCGAACACCAATTTGATGCTTTAAGA 26
|||||
Db 686 GCGATCTCAAAATTTGATGCTTTAAGA 711

RESULT 8
US-10-027-632-174961/c
; Sequence 174961, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 174961
; LENGTH: 3186778
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1) ... (3186778)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-174961
Query Match 73.8%; Score 19.2; DB 5; Length 3186778;
Best Local Similarity 87.5%; Pred. No. 1.2e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 GAACACCAATTTGATGCTTTAAGA 26

```
|||||
Db 2232992 GAACACCAATTAGATGCTTGGAGA 2232969

RESULT 9
US-10-027-632-174961/c
; Sequence 174961, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 174961
; LENGTH: 3186778
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(3186778)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-174961

Query Match 73.8%; Score 19.2; DB 6; Length 3186778;
Best Local Similarity 87.5%; Pred. No. 1.2e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 GAACACCAATTGATGCTTTAAGA 26
|||||
Db 2232992 GAACACCAATTAGATGCTTGGAGA 2232969

RESULT 10
US-09-925-065A-593526/c
; Sequence 593526, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; NUCLEOTIDE POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 4328
; TYPE: DNA

Query Match 72.3%; Score 18.8; DB 4; Length 578;
Best Local Similarity 90.9%; Pred. No. 4.1e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 ACACCAATTGATGCTTTAAGA 26
|||||
Db 436 ACACCAATTGATGCTTTAAGA 415

RESULT 11
US-09-925-065A-722502/c
; Sequence 722502, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; NUCLEOTIDE POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 722502
; LENGTH: 603
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-722502

Query Match 72.3%; Score 18.8; DB 4; Length 603;
Best Local Similarity 90.9%; Pred. No. 4.2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 ACACCAATTGATGCTTTAAGA 26
|||||
Db 94 ACACCAATTGATGCTTTAAGA 73

RESULT 12
US-09-908-299-4/c
; Sequence 4, Application US/09908299
; Publication No. US20030051273A1
; GENERAL INFORMATION:
; APPLICANT: MARY WILDERMUTH,
; APPLICANT: JULIA DEWDNEY
; TITLE OF INVENTION: SALICYLIC ACID BIOSYNTHETIC GENES AND
; USES THEREOF
; FILE REFERENCE: 00786/391002
; CURRENT APPLICATION NUMBER: US/09/908,299
; CURRENT FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/219,231
; PRIOR FILING DATE: 2000-07-18
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 4328
; TYPE: DNA
```

```
; ORGANISM: Arabidopsis thaliana
US-09-908-299-4

Query Match      72.3%; Score 18.8; DB 3; Length 4328;
Best Local Similarity 90.9%; Pred. No. 6.2e+02;
Matches 20; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      4 AACACCAATTGATGCTTTAAG 25
      |||||
Db      250 AACACCAATTGATGCTTTAAG 229

RESULT 13
US-09-925-065A-679116
; Sequence 679116, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 679116
; LENGTH: 656
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-679116

Query Match      70.8%; Score 18.4; DB 4; Length 656;
Best Local Similarity 95.0%; Pred. No. 6.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 AACACCAATTGATGCTTTA 23
      |||||
Db      127 AACACCAATTGATGCTTTA 146

RESULT 14
US-09-925-065A-527754/c
; Sequence 527754, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 527754
; LENGTH: 548
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-527755/c
; Sequence 527755, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 527755
; LENGTH: 548
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-527755

Query Match      70.0%; Score 18.2; DB 4; Length 548;
Best Local Similarity 87.0%; Pred. No. 7.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      3 GAACACCAATTGATGCTTTAAG 25
      |||||
Db      322 GAGCACCAATATGATGCTTTAAG 300

Search completed: November 27, 2005, 02:22:42
Job time : 37.9737 secs
```

This Page Blank (uspic)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:41:08 ; Search time 9.17647 Seconds
(without alignments)
421.056 Million cell updates/sec

Title: US-09-555-529-19

Perfect score: 26

Sequence: 1 gcaacacacaaatttgatgctttaaga 26

Scoring table:

IDENTITY|NUC
Gapop 10|0 , Gapext 1.0

Searched: 3205263 seqs, 74304013 residues

Total number of hits satisfying chosen parameters: 6410526

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA New:

1: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
2: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq2.*
9: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq3.*
10: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	16.8	64.6	1688	1	US-10-510-386-157
C 2	16.2	62.3	423	1	US-10-793-626-2609
C 3	16.2	62.3	579	1	US-10-793-626-1999
C 4	16.2	62.3	579	1	US-10-793-626-2303
C 5	16.2	62.3	1809	1	US-10-793-626-2361
C 6	16.2	62.3	2085	1	US-10-485-517-391
C 7	16.2	62.3	2876	1	US-10-793-626-3778
C 8	16.2	62.3	3169	1	US-10-793-626-3865
C 9	16.2	62.3	3591	1	US-10-793-626-4033
C 10	16.2	62.3	3627	1	US-10-793-626-4311
C 11	16.2	62.3	4116	1	US-10-793-626-4127
C 12	16.2	62.3	40439	1	US-10-993-509-1
C 13	16.1	61.5	1948	1	US-10-510-386-243
C 14	16.1	61.5	3069	1	US-10-793-626-4068
C 15	15.8	60.8	19	9	US-11-101-244-36671
C 16	15.8	60.8	19	9	US-11-083-784-36671
C 17	15.6	60.0	1118	9	US-11-082-389-131
C 18	15.6	60.0	1560	9	US-11-082-389-129
C 19	15.6	60.0	1929	9	US-11-074-176-317
C 20	15.6	60.0	2001	7	US-11-074-176-63
C 21	15.6	60.0	2409	1	US-10-689-742-73
C 22	15.6	60.0	2734	1	US-10-510-386-5
C 23	15.6	60.0	7117	7	US-11-082-544-34

C 24	15.4	59.2	3059	1	US-10-793-626-3651	Sequence 3651, Ap
C 25	15.2	58.5	450	1	US-10-793-626-351	Sequence 351, App
C 26	15.2	58.5	789	1	US-10-793-626-2029	Sequence 2029, Ap
C 27	15.2	58.5	1829	1	US-10-667-295-200	Sequence 200, App
C 28	15.2	58.5	2259	1	US-10-793-626-347	Sequence 347, App
C 29	15.2	58.5	2582	1	US-10-485-517-35	Sequence 35, Appl
C 30	15.2	58.5	2582	1	US-10-485-517-105	Sequence 105, App
C 31	15.2	58.5	3007	1	US-10-485-517-34	Sequence 34, Appl
C 32	15.2	58.5	3038	1	US-10-793-626-3383	Sequence 3383, Ap
C 33	15.2	58.5	3328	1	US-10-793-626-3446	Sequence 3446, Ap
C 34	15.2	58.5	3328	1	US-10-793-626-3644	Sequence 3644, Ap
C 35	15.2	58.5	3701	1	US-10-793-626-3392	Sequence 3392, Ap
C 36	15.2	58.5	7944	1	US-10-821-234-451	Sequence 451, App
C 37	15.2	58.5	340000	7	US-11-102-978-3	Sequence 3, Appli
C 38	15	57.7	2865	1	US-10-467-962B-30	Sequence 30, Appli
C 39	14.8	56.9	19	8	US-11-101-244-36654	Sequence 36654, A
C 40	14.8	56.9	19	8	US-11-101-244-36679	Sequence 36679, A
C 41	14.8	56.9	19	8	US-11-101-244-605665	Sequence 605665,
C 42	14.8	56.9	19	8	US-11-101-244-1426752	Sequence 1426752,
C 43	14.8	56.9	19	9	US-11-083-784-36654	Sequence 36654, A
C 44	14.8	56.9	19	9	US-11-083-784-36679	Sequence 36679, A
C 45	14.8	56.9	19	9	US-11-083-784-605665	Sequence 605665,

ALIGNMENTS

RESULT 1
US-10-510-386-157/c
; Sequence 157, Application US/10510386
; Publication No. US20050244922A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Jens Tonne
; APPLICANT: Clausen, Ib Groth
; APPLICANT: Jorgensen, Steen Troels
; APPLICANT: Olsen, Peter Bjarke
; APPLICANT: Rasmussen, Michael Dolberg
; TITLE OF INVENTION: Improved Bacillus Host Cell
; FILE REFERENCE: 10294, 204-US
; CURRENT APPLICATION NUMBER: US/10/510,386
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 248
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 157
; LENGTH: 1688
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (501)..(1652)
US-10-510-386-157

Query Match 64.6%; Score 16.8; DB 1; Length 1688;
Best Local Similarity 90.0%; Pred. No. 38;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 GAACACCAATTGATGCTTT 22
|||||||
Db 443 GAACACCAATTGATGATTT 424

RESULT 2
US-10-793-626-2609
; Sequence 2609, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09

```

; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2609
; LENGTH: 423
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-2609

Query Match          62.3%; Score 16.2; DB 1; Length 423;
Best Local Similarity 85.7%; Pred. No. 54; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      6 CACCAATTTGATGCCTTTAAGA 26
Db       ||||||| ||||| |||||
        94 CATCAATTTGATGATTTAAAA 74

RESULT 3
US-10-793-626-1999/c
; Sequence 1999, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; PRIOR FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1999
; LENGTH: 579
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-1999

Query Match          62.3%; Score 16.2; DB 1; Length 579;
Best Local Similarity 85.7%; Pred. No. 58; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      6 CACCAATTTGATGCCTTTAAGA 26
Db       ||||||| ||||| |||||
        94 CATCAATTTGATGATTTAAAA 74

RESULT 4
US-10-793-626-2303/c
; Sequence 2303, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; PRIOR FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2303
; LENGTH: 579
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-2303

Query Match          62.3%; Score 16.2; DB 1; Length 579;
Best Local Similarity 85.7%; Pred. No. 58; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      6 CACCAATTTGATGCCTTTAAGA 26
Db       ||||||| ||||| |||||
        94 CATCAATTTGATGATTTAAAA 74

RESULT 5
US-10-793-626-2361
; Sequence 2361, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; PRIOR FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2361
; LENGTH: 1809
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-2361

Query Match          62.3%; Score 16.2; DB 1; Length 1809;
Best Local Similarity 85.7%; Pred. No. 74; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 CGAACACCAATTGTGCTTTT 22
Db       ||||||| ||||| |||||
        1767 CGAACAAAAATTAGATGCTTT 1787

RESULT 6
US-10-485-517-391
; Sequence 391, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WO
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 391
; LENGTH: 2085
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-10-485-517-391

Query Match          62.3%; Score 16.2; DB 1; Length 2085;
Best Local Similarity 85.7%; Pred. No. 76; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      6 CACCAATTTGATGCCTTTAAGA 26

```


RESULT 9
US-10-793-626-4033/c
; Sequence 4033, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:

```

1  GENERAL INFORMATION:
2  APPLICANT: KIMBERLY, WILLIAM JOHN
3  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
4  FILE REFERENCE: PU3480US
5  CURRENT APPLICATION NUMBER: US/10/793,626
6  CURRENT FILING DATE: 2004-03-04
7  PRIOR APPLICATION NUMBER: 60/164,258
8  PRIOR FILING DATE: 1999-11-09
9  NUMBER OF SEQ ID NOS: 4472
10 SOFTWARE: Patent In. Ver. 2.1

```


Db 1 AACACUAAUUGCUGCUUU 19

Search completed: November 27, 2005, 02:44:31
Job time : 10.1765 secs

Best Local Similarity 79.2%; Pred. No. 93;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

Qy  2  CGAACACCAATTTGATGCTTTAAG  25
      |||||
Db 792 CAATCACCGATTTAATGCTGTAAG  769
      |||||

```

Search completed: November 27, 2005, 02:44:31
 Job time : 10.1765 sec

```

RESULT 14
US-10-793-626-4068/c
; Sequence 4068, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4068
; LENGTH: 3069
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4068

```

Query Match	61.5%;	Score 16;	DB 1;	Length 3069;
Best Local Similarity	79.2%;	Pred. No. 1e+02;		
Matches 19:	Conservative	0;	Mismatches	5;
			Indels	0;
			Gaps	0;

Qy 3 GAACACCAATTGTGCTTAAAGA 26
||||| ||||| | |||||
pb 3020 GAACAACCAATTGTGAAGAATTAAA 2997

```

RESULT 15
US-11-101-244-36671
; Sequence 36671, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 36671
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-36671

```

Query Match	60.8%	Score 15.8;	DB 8;	Length 19;
Best Local Similarity	52.6%;	Pred. No. 44;		
Matches	10. Conservative	7: Mismatches	2: Indels	0: Gaps

QY 4 AACACCAATTGTATGCTTT 22

This Page Blank (uspto)

C:Superfamily: caldesmon
C:Keywords: actin binding; calmodulin binding; phosphoprotein
F:342,427,462/Binding site: phosphate (Ser) (covalent) #status predicted
F:433,456/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 7.2%; Score 145.5; DB 1; Length 517;
Best Local Similarity 23.6%; Pred. No. 0.16; Indels 83; Gaps 12;
Matches 68; Conservative 46; Mismatches 91; Indels 83; Gaps 12;

Qy 48 ESHQRQLLASNPQFMDYFSEEFNRDFLELLRRFGTKRVHNNIVYNEYISHREHIHM 107
Db 90 ERRQKRLQALERQKEFDFTIDGSL-----VPSRRVNVNVEINEITGKEKV-- 138

Qy 108 NATQWETITDFTKWLGRGLCKVDETPKGWYQYIDRDPETIR-----RQLELEKKK 159
Db 139 -----ET-----RQRCBIET-----ETVTKSYQRNMWRQDGEEGK 171

Qy 160 KODLDEEKTAFIEEQVRRGLEGEQETPVFTLSRENEBEKVTFNLNKA-----GG 213
Db 172 KEEKDEEKEPKVPTENQVKNKVK-----KAPKEEMKSVNDRKRGVPEQKAQNG 224

Qy 214 SAGATTSK-----SSSLGPSALKLGSASG--KRKSSQSSA---QPAKKKKKLSALDEIME 264
Db 225 ERELTPKLKSTENAFGRSLNKAANAAGSEKLKEQQAVALDELKCRREERRKILE 284

Qy 265 LBEKKQRTARTAWLPQIGIVVKIITKJLGEKHKKKGVVKEVIDRYTA 312
Db 285 EEEQKKQFEAE-----RKIREBEKKR--MKEBIERRA 317

RESULT 8
H96770
protein heat shock protein F1017.8 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: H96770
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A66141; MUID:21016719; PMID:11130712
A:Accession: H96770
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-630 <STO>
A:Cross-references: UNIPROT:Q9C911; UNIPARC:UPI00000AA4E1; GB:AE0051173; NID:G6939229; PI
C:Genetics:
A:Gene: F1017.8
A:Map position: 1

Query Match 6.6%; Score 134.5; DB 2; Length 630;
Best Local Similarity 20.9%; Pred. No. 0.96;
Matches 73; Conservative 65; Mismatches 121; Indels 91; Gaps 15;

Qy 27 YCOMQKQCRDENGPKCHMSHQRQLLASNPQFMDYFSEEFNRDFLELLRRFGT 86
Db 309 YCIVCSKFKSEKQKNHEQSKGKKEV---AELRESFTDYEEENEEDIDGLDPSPEV 365

Qy 87 KRHHNNIVNEYISHREHIHMNATQWETITDFTKWLGRGLCKVDETPKGWYI----- 139
Db 366 BELHKLQELNIDNE-----RDVKVGVGEADETDDEYFVASEDMQG 409

Qy 140 --QYIDRPE--TIRQLELEKKKKQ-----DLDEEKTAKFIE-----EQVRR 179
Db 410 SSESEDEDDMTLLKWKVSGQKNQKNNVSVKBEDETEVEIEGDTAEFSEFDNKGSTGR 469

Qy 180 GLEGKEQETPVFTLSRENEBEKVTFNLN-----GNAG-----SAGATTSKSSSLGPS 228

Db 470 NKEAKE-----ERNKONAGNDVADDTSKVQIPGEGGNPDENMMNATESASGALADSQKD 522
Qy 229 ---ALKLGSAAAGKRKES-----SQSSAPAKKKKKSALDEI-----MELEBEKK--RTA 273
Db 523 EANSMEYDNRKSTGRRRSKKGDKNNQGLNEKSSEADDTQYVNRDMESQDYKKAPRSK 582

Qy 274 RTDAWLQPGIVVKIITKK-----LGEKYHKKKGVVKEVIDRYTAVVK 315
Db 583 KSTR-----GMKTKGTGTTKNSNECDRCGEFESRTKLJHKLHADSGHATVK 628

RESULT 9
T29340
hypotheetical protein F21C10.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T29340
R:Du, Z.; Gattung, S.
submitted to the EMBL Data Library, April 1996
A:Description: The sequence of C. elegans cosmid F21C10.
A:Reference number: Z20610
A:Accession: T29340
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2541 <DUZ>
A:Cross-references: UNIPROT:Q19663; UNIPARC:UPI0000076CB3; EMBL:U55364; PIDN:AAA97973.1;
A:Experimental source: strain Bristol N2; clone F21C10
C:Genetics:
A:Gene: CESP:F21C10.7
A:Map position: 5
A:Introns: 50/3; 91/3; 153/3; 184/3; 215/1; 254/3; 350/1; 392/3; 538/1; 589/3; 625/3; 751/3

Query Match 6.5%; Score 132.5; DB 2; Length 2541;
Best Local Similarity 21.8%; Pred. No. 6.9;
Matches 72; Conservative 55; Mismatches 124; Indels 79; Gaps 11;

Qy 97 EYISHREHHNATQWETITDFTKWLGRGLCKVDETPKGW-----YIQVIDRDP--ET 148
Db 852 EFVHKMEVEHVQEQKLKLTVELEBEKLEQQLVKVTEHIQMWQEMIEIIRKMFDPSTPMKT 911

Qy 149 IRRQLELEKKKKQDLDDDEKTAFTTEOVRRGLEGEQETPVFTLSRENEBEKV----- 203
Db 912 VOESQELQEKVKLKEALEVQPKIEDV-----AKADIAVKTQISRAVEKQVIRENA 966

Qy 204 -----TFNLNKGAGSAGATTSSSLGPSALKLGSAAAGKRKESQSSAQPAK 253
Db 967 ETLEKKAISATFELSKPQQTIEETKTILTSL-----EQQTILTPEE 1009

Qy 254 KKSALDIMELEBEKKKTARTDAWLQPGI--VVKII--TKKLGKVKHKKGVVKEVIDRY 310
Db 1010 KQ-----ELQILKKTVEEIQMWQETIEIQLVDKTPKTVQSESTKIKKKNBIQQTV 1061

Qy 311 TAVVKMTSDGRLKLDQTHLETVPAPGKRVLVNGGYRGNEGTLSEINEKAFSATIVIE 370
Db 1062 VAQTRLEDSARFRKDEFTTKTVQETWTKQQQV-----QQLVKELHERAASQNVIVR 1113

Qy 371 TGPLKGRV-----EGIQYE 385
Db 1114 QKEERSRVQAPQIITQLKDDVDVDEGCRVE 1143

RESULT 10
T05523
hypotheetical protein F13M23.160 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004
C:Accession: T05523
R:Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hoheisel, J.; Mewes, H.W.; Mayer, K.F.;
submitted to the Protein Sequence Database, February 1999
A:Reference number: T15419
A:Accession: T05523
A:Molecule type: DNA

A:Residues: 1-375 <BEV>
A:Cross-references: UNIPROT:Q5SW24; UNIPARC:UPI00000A6P0F; EMBL:AL035523
A:Experimental source: cultivar Columbia; BAC clone F13M23
C:Genetics:
A:Map position: 4
A:Note: F13M23.160

Query Match 6.5%; Score 131; DB 2; Length 375;
Best Local Similarity 23.4%; Pred. No. 0.83;
Matches 62; Conservative 51; Mismatches 102; Indels 50; Gaps 12;

QY 132 ETPKGVYQVDRDPETIRRLQLEKKKKQDLDEEKTAKFIEQVRRGLEG-----KE 185
DB 80 ETPSG-----DGSNNIYGLNL-RQVENVEKPEVEKMLKSMRTLESIPDAPEKE 132
QY 186 --QETPV--FTE-----LSRENEEEKVTNMLKAGSGAGATTSSSLG 226
DB 133 DFESFVDGFEALLAGYWKPGQIGLKAKEDVKIVEKWSGNEGFGFGKSDKAMKMI 192
QY 227 PSALKLLGSAS-----GKKESSQSAQPAKKKKSALD-----EIMELEE-----EKKRT 272
DB 193 DN--KLVGSGSHEVKVGINKIENMKDRVVRKRNRETEGESTEVKAACKQNTRGQTRET 250
QY 273 ARTDAMLQPGIVVKIITKKL-GEKYHKKGKGVKEVIDRYTAVVVKMTDSDGRLK-LDQTHL 330
DB 251 REKTSMLRSHIKVRIISKDVKGRLYLKKAVIDDVVGPTSCDIAMDETQELVQGIQELL 310
QY 331 ETVIPAPGKRVLVINGYRQNEGTL 355
DB 311 ETALPRGGSVLVLSGRHKGVYGRLL 335

RESULT 11
T24518
hypothetical protein T05D4.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T24518
R:McMurray, A.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19902
A:Accession: T24518
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-482 <WIL>
A:Cross-references: UNIPROT:O45751; UNIPARC:UPI000007C558; EMBL:Z81115; PIDN: CAB03295.1;
A:Experimental source: clone T05D4
C:Genetics:
A:Gene: CESP:T05D4.5
A:Map position: 3
A:Introns: 128/3; 184/3; 212/3; 236/3; 305/3

Query Match 6.4%; Score 130.5; DB 2; Length 482;
Best Local Similarity 21.0%; Pred. No. 1.2;
Matches 78; Conservative 55; Mismatches 146; Indels 93; Gaps 15;

QY 2 GKSDFLSPK---AJANRIKSGLQKLRWYC-----QMCQOCQRENGCFKCH----- 44
DB 118 GKCPYFDQRLITVGRVGEVPVQ--HWFAIYNTSPFPYMFKIKCSNLSLFKHPVCGVL 175
QY 45 -CMSESHQRLQLASENPQPMDFYS-----EPPNDLFELLRLRFEGTKRV 89
DB 176 DALCDABISVTFVGAHVPPRYTSYLSIHSMLPDPFLDVQAFRN-----RRNFETRQ- 228
QY 90 HNNIVYNEYISHREHIHNNATQWETLTDFTKWLREGLCCKVDETPKGYIYQIDR----- 144
DB 229 ---FLYVEYAHEDPI-----LVTAEDRDQFHFQYMEKVLKTD 264
QY 145 DPETIRRLQLEJEXKKKQDLDD-DESKTAKFIEQVRRGLEKQBQETPVPTELSENREBEKV 203
DB 265 EAESFAPKLLAEKVERLDKKIDEIKARKIDKLTIER--EIVRAESPKEPRTLRVKEKF 322
QY 204 TPNLNGKAGGS-----AGATTSSSSLGPSALKLLGSNASGKR-----KESSQSAQPAKK 254

DB 323 -FNQKQKKNENWMEKLVDAASRKPMLSHWDLKMKIKAEKKKIGAVKKKIKAEKKEDKK 381
QY 255 KKSALDEIMELEBEKKKRTARTDAMLQPGIVVKIITKKLGEKYHKKGKGVVKEVIDRYTAVV 314
DB 382 KPMGMNSTLJMERDERRTKRKEE-----EKKKKDAEKKKKDAEKKKDAEDE 431
QY 315 KMTDSGDRLLKLD 326
DB 432 KKKDAGDEKKKD 443

RESULT 12
S23325

M2 protein precursor - Streptococcus pyogenes
C:Species: Streptococcus pyogenes
A:Variety: serotype M2
C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C:Accession: S23325; S35761; S61078; S60785
R:Bessen, D.E.; Fischetti, V.A.
Infect. Immun. 60, 124-135, 1992
A:Title: Nucleotide sequences of two adjacent M or M-like protein genes of group A streptococci
A:Reference number: S23325; MUID: 92104662; PMID: 1370269
A:Accession: S23325
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-407 <BES>
A:Cross-references: UNIPROT:P50468; UNIPARC:UPI000012EABE; EMBL:X61276; NID:G47369; PIDN:
R:Podbielski, A.
submitted to the EMBL Data Library, November 1992
A:Reference number: S35760
A:Accession: S35761
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-17 <POD>
A:Cross-references: UNIPARC:UPI00017823B; EMBL:X69324
R:Whitmore, A.; Kapur, V.; Sullivan, D.; Musser, J.; Kehoe, M.
submitted to the EMBL Data Library, July 1994
A:Description: Noncongruent relationships between variation in emm gene sequences and the
A:Reference number: S60784; MUID: 95198537; PMID: 7891551
A:Accession: S60785
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 31-89 <WH2>
A:Cross-references: UNIPARC:UPI000017823C; EMBL:U11958
C:Superfamily: M5 protein

Query Match 6.3%; Score 128.5; DB 2; Length 407;
Best Local Similarity 21.0%; Pred. No. 1.3;
Matches 71; Conservative 69; Mismatches 149; Indels 49; Gaps 11;

QY 70 EEFNDPFLLELRRRFGTKRVHNNIVYNEYISHRE-HIHMNATQWETLTDFTKWLREGLC 128
DB 90 EEKHKDEKLEKSESDVERHYLRQLDQEYEQEQQRQKLELELRQSQREVSKRY-QEQLQ 148
QY 129 KVDETPKGVYQVDRDPETIRRLQLEKKKKQDLDD-----DEETAKFIEQVTRGL 181
DB 149 KQOQLEKEKQISEASR--KSLRRDLASRAAKKDLAEHQKLEKEKQISEASRSKSLRDL 206
QY 182 E-GKEQETPVPTELSENREBEKVTFNMLKAGSGAGATTSSSSLGPSALKL-----LGS 235
DB 207 EAGRAAKKDLAEHQKLEKEKQISEASRQGLSRDLASRAAKKDLAEHQKLEKEKQISE 266
QY 236 A-----ASGKRKESSQSAQPAKKKKKSALDEI-MELEBEKKRTARTDAMLQPGIV 284

Db 267 ASRQGLSRDLASRAKKVVEADLAEANSKLQALEKINKLEBEGKKLSKEKAELOAKLE 326
QY 285 VKIITKKLGEYHKKKKVGVVKEVIDRYTAVVVKMTDSDGRKLQDQTHLETVIPAGKRVILV 344
Db 327 AE--AKALKEQLAKQAEELAKLKGNOTPNAKVAPOANRSRSAMTQKRTLPTSG----- 378
QY 345 NGYRGNEGTLSINE--KAFSATIVETGPKGRARVE 380
Db 379 -----ETANPFPTAAATAATVMSAGMLAKRKE 405
RESULT 13
A84671
hypothetical protein At2g27280 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C:Accession: A84671
R.;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.B.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; PMID:20083487; PMID:10617197
A:Accession: A84671
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-764 <STO>
A:Cross-references: UNIPROT:Q9XKNS; UNIPARC:UPI000017A6B9; GB:AE002093; NID:G5306269; PI
A:Gene: At2g27280
A:Map position: 2
Query Match 6.3%; Score 128.5; DB 2; Length 764;
Best Local Similarity 21.0%; Pred. No. 2.8;
Matches 74; Conservative 69; Mismatches 116; Indels 93; Gaps 16;
QY 46 MESHQRQLLASNPQQF-----MDYFSEFPNDFLELLRRFGTK 87
Db 54 IEEOHKKAL---EEDPSAFSYDEVYDDMKQKAVLPRMQD--REERKPRYIQNLKMQAER 108
QY 88 RVHNNIVNVEYISHR-----EHTHMNATQWET-----LTDFTKWLGRGLCKVDPTPKW 137
Db 109 EKEHEIVRYERKLAKREKDEHLFSDKKEFTVGYAKRKLEEQKKWLAEEKLRELR----- 163
QY 138 YIQYIDRDDETIRRLQ-----ELEKKKKQDLDDEBKTAKFTIEEQVRRGL 181
Db 164 -----ERDDVTKKKOLSDFYFNIGNVAFGAREVEAKEAEKLEEQKAEKLEEQKAEKL 218
QY 182 EKEQETPVFTLSRENEBEKVTFNINLKAGGSAGATTSSKSSSLGP-----SALKLGS 235
Db 219 EELRKEV-TRVEKGRKSPEKEV--SPDSGEFGS-----SRKSLSLEPLEAEQAVSEKMG 270
QY 236 AASGRKSSQSSQAQPAK-----KKSALDEIMELEBEKKRTARTDAWLQPGIVVKLIITKK 291
Db 271 DQTEERKSSIKAEKAVKPAINDQKRDEAIAAAKE--RFLARST-----ITNA 318
QY 292 LGEKYHKKGVVKEV--IDRYTAVVKMTDSDGRKL--LDQTHLETVIPAGK 339
Db 319 IGLVKNKEENQKIPYVEMHIVRVSINVLVQIEWYICKRFLNVLDFIGK 370
RESULT 14
T34036
hypothetical protein B0041.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C:Accession: T34036
R.;Fulton, R.; Wohlmann, P.
submitted to the EMBL Data Library, April 1997
A:Description: The sequence of C. elegans cosmid B0041.
A:Reference number: Z21466
A:Accession: T34036

A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1359 <FUL>
A:Cross-references: UNIPROT:Q9U7E0; UNIPARC:UPI0000060B63; EMBL:AF000196; PIDN:AAC24256.3
A:Experimental source: strain Bristol N2; clone B0041
C:Genetics:
A:Gene: CESP.B0041.7
A:Map position: 1
A:introns: 12/2; 59/2; 248/2; 582/1; 675/1; 733/3; 966/2; 1044/2; 1310/3
Query Match 6.3%; Score 127; DB 2; Length 1359;
Best Local Similarity 22.3%; Pred. No. 7;
Matches 63; Conservative 47; Mismatches 107; Indels 66; Gaps 9;
QY 145 DPETIRROLELEKK--KKQDLDDEBKTAKFTIEEQVRGLEKEQETPVFTLSRNEBEK 202
Db 214 DEKEVKSKKKSKVVKKESEDEDAPEKKTKTKRKTSSSESE--SEKSDSEEEK 271
QY 203 VTFNLKAGGSAGATTSSSSLSALKLGSASGKRKSSQSSAOPAKKKKSALDEI 262
Db 272 -----ESSPPKPKKKPLAVKKL----SSDESEESDVEVLPQKKRGAVTLI 314
QY 263 MELEBEKKRTARTDA-----WLQPGIVVKIITKKLGEYHKK 299
Db 315 SDSEDEKQKSESEASDYEEKVKKKAKQSESSGSDSSSGSITVNRKSKKKEPEKKK 374
QY 300 KGVV-----KEVIDRYTAVVKMTDSDGRKLQDQTHLETVIPAGKRVILVNGYRGNE 352
Db 375 KGIIMDSKLOKETIDAERA---EKERRKRLEKKQKEFNGIVLBERGEDILTEMLTG--TSSQ 430
QY 353 GTLESINEKAFSATIVETGP-----LKGRRVEGQY 384
Db 431 RKLKSVLDPDSSTVDEESKXPVEVHNSLVRLKPKHQAHGQF 473
RESULT 15
A26655
myosin heavy chain [similarity] - slime mold (Dictyostelium discoideum)
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Dictyostelium discoideum
C:Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 09-Jul-2004
C:Accession: A26655; A24728; S00250
R.;Warrick, H.M.; De Lozanne, A.; Leinwand, L.A.; Spudich, J.A.
Proc. Natl. Acad. Sci. U.S.A. 83, 9433-9437, 1986
A:Title: Conserved protein domains in a myosin heavy chain gene from Dictyostelium discoi
A:Reference number: A26655; MUID:87092266; PMID:3540939
A:Accession: A26655
A:Molecule type: DNA
A:Residues: 1-2116 <WAR>
A:Cross-references: UNIPROT:P08799; UNIPARC:UPI000012FBD3; GB:M14628; GB:M11938; NID:G16;
R.;DeLozanne, A.; Lewis, M.; Spudich, J.A.; Leinwand, L.A.
Proc. Natl. Acad. Sci. U.S.A. 82, 6807-6810, 1985
A:Reference number: A24728; MUID:86016788; PMID:3901008
A:Accession: A24728
A:Molecule type: mRNA
A:Residues: 2035-2116
A:Cross-references: UNIPARC:UPI0000177634
R.;Wagle, G.; Noegel, A.; Scheel, J.; Gerisch, G.
FEBS Lett. 227, 71-75, 1988
A:Title: Phosphorylation of threonine residues on cloned fragments of the Dictyostelium n
A:Reference number: S00250; MUID:88112226; PMID:2828113
A:Accession: S00250
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1734-1893 <WAG>
A:Cross-references: UNIPARC:UPI0000177635
C:Comment: The rod domain is highly periodic, containing a pattern of 7-residue repeats
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; hydroxylase; nucleotide binding; P-loop; phosph
F.1-818/Domain: globular head <HED>
F.89-747/Domain: myosin motor domain homology <MMOT>
F.179-186/Region: nucleotide-binding motif A (P-loop)
F.819-2116/Domain: alpha-helical rod <ROD>

```
Query Match      6.3%; Score 127; DB 2; Length 2116;
Best Local Similarity 27.0%; Pred. No. 12; Mismatches 72; Indels 64; Gaps 11;
Matches 64; Conservative 37;

QY 129 KVDETPKGWYIQYIDRDPETIRRLQLELE-----KQKKQDLDDBEKTAKFIEEQV 177
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1171 KIKSTLEG-----EVARLQGELEAEQAKSNVEKQKKVLELDEKSAQLAEETA 1220
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 178 RRGLEOK-----EQE-TPVFTLSRENEEEKVTFNLNKGAGGSAGATTSSSSLGPSALK 231
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1221 AKQALDKLKKLEQELSEVQTLSEANNK-----NYSND-----STNKHLETSEFNNLK 1268
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 232 LLGSAASGKRRKSSOSSAOPAKKKKSDALDEIM-----ELEEEKKRTARTD----- 276
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1269 L-----ELEAEQAKAKQALEKKRGLGSELSKHVNEQLEEEKKQKESNEKRVKVDLEKEV 1320
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 277 AWLQPGIVVKIITPK-LGEKYHKKGWKEVIDRYTAVVVKMTD-SGDRKLKLDQTHLE 331
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1321 SELKDQIEEVASKKAVTEAKNKKSESELDEIKQYADVSSRDKSVQELKTLQAKNE 1377
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: November 23, 2005, 16:34:28
Job time : 27.9362 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:06:33 ; Search time 130.167 Seconds
(without alignments)
1319.820 Million cell updates/sec

Title: US-09-555-529-25
Perfect score: 2028
Sequence: 1 MGKSDPLSPKATANRIKSG.....GPKGRVGEIQVEDISKLA 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_21.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*
9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2020	99.6	391	2	AAY23619 Murine ki
2	1895	93.4	393	2	AAY23620 A human k
3	1895	93.4	393	7	ADJ69659 Human hea
4	1456	71.8	291	2	AAY23617 Mouse tru
5	1365	67.3	293	2	AAY23618 A huamn t
6	1245	61.4	390	8	ADP22452 Sea-squir
7	1052.5	51.9	1390	4	ABE71416 Drosophil
8	988	48.7	411	3	AG42576 Arabidops
9	913	45.0	424	3	AB03063 Maize KIN
10	913	45.0	437	8	ADX96318 Arabidops
11	913	45.0	437	8	ADX78817 Plant ful
12	901.5	44.5	423	3	AB03064 Maize KIN
13	901.5	44.5	423	3	AB03065 Maize KIN
14	901.5	44.5	423	8	ADT58422 Plant pol
15	843	41.6	382	3	AG42577 Arabidops
16	740	36.5	366	3	AG42578 Arabidops
17	607	29.9	340	3	AG46711 Arabidops
18	607	29.9	343	3	AG46710 Arabidops
19	590	29.1	302	3	AG46712 Arabidops
20	361	17.8	84	4	AG474690 Human col
21	183	9.0	82	3	AG19349 Arabidops
22	170	8.4	81	3	AG19350 Arabidops
23	162	8.0	76	3	AG19351 Arabidops
24	152	7.5	76	3	AG46729 Arabidops

Adx77206 Plant ful
Aar22904 1-Caldeem
Aag21445 Arabidops
Adt56172 Plant pol
Aag43126 Arabidops
Aag50080 Arabidops
Abp97029 Rat L-FIL
Aag85008 Shrimp wh
Aag43128 Arabidops
Aag21447 Arabidops
Aag21446 Arabidops
Aag43127 Arabidops
Aag50082 Arabidops
Aag50081 Arabidops
Adp49328 S pyrogen
Aag15386 Arabidops
Ad823527 Bacterial
Aag15388 Arabidops
Aag15387 Arabidops
Abp96033 Human sur
Aao16359 Human tra

25 149 7.3 278 8 ADX77206
26 142.5 7.0 517 2 AAR22904
27 134 6.6 375 3 AAG21445
28 134 6.6 375 8 ADT56172
29 131 6.5 375 3 AAG43126
30 131 6.5 450 3 AAG50080
31 130.5 6.4 1212 6 ABP97029
32 129.5 6.4 1141 4 AAG85008
33 128.5 6.3 258 3 AAG43128
34 128.5 6.3 258 3 AAG21447
35 128.5 6.3 263 3 AAG21446
36 128.5 6.3 263 3 AAG43127
37 128.5 6.3 333 3 AAG50082
38 128.5 6.3 338 3 AAG50081
39 128.5 6.3 407 8 ADP49328
40 125.5 6.2 323 3 AAG15386
41 124.5 6.1 1080 8 ADS23527
42 124 6.1 237 3 AAG15388
43 124 6.1 246 3 AAG15387
44 124 6.1 726 6 ABP96033
45 124 6.1 2349 6 AAO16359

ALIGNMENTS

RESULT 1
AAY23619
ID AAY23619 standard; protein; 391 AA.
XX
AC AAY23619;
XX
DT 07-SEP-1999 (first entry)
XX
DE Murine kin17 protein.
XX
KW Mouse; kin17 protein; cell proliferation; fertility;
KW hyperproliferative disease; protein interaction; curved DNA;
KW HIV replication; HIV integration; repair enzyme.
XX
OS Mus sp.
XX
PN FR2772046-AL.
XX
PD 11-JUN-1999.
XX
PF 09-DEC-1997; 97FR-00015536.
XX
PR 09-DEC-1997; 97FR-00015536.
XX
(COMS) COMMISSARIAT ENERGIE ATOMIQUE.
XX
Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
WPI; 1999-359999/31.
XX
New DNA coding for human kin17 protein - useful for controlling cell proliferation or fertility.
XX
Claim 19; Page 36-37; 69pp; French.
XX
The present sequence represents a murine kin17 protein. The mammalian kin17 protein is useful for preparing a medicament for controlling cell proliferation or for controlling fertility. The medicaments can also be used to treat hyperproliferative diseases. Fragments between amino acids 55 and 235 (preferably between amino acids 129 and 228) of a mammalian kin17 protein are useful for regulating the interaction between proteins and curved DNA. The fragment can be used to block replication of HIV or its integration into the human genome or to target repair enzymes to curved DNA sites. Expression vectors for kin17 can be used for controlling cell proliferation
XX
Sequence 391 AA;
SQ

Query Match	99.6%	Score 2020;	DB 2;	Length 391;
Best Local Similarity	99.7%;	Pred. No. 1.4e-170;		
Matches 390; Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;	

Qy	1	MGKSDFLSPKAIANRIKSGIQKLRYWCQMCQKCRDENGFKCHCWSHSHORQLLASEN	60
Db	1	MGKSDFLSPKAIANRIKSGIQKLRYWCQMCQKCRDENGFKCHCWSHSHORQLLASEN	60
Qy	61	PQOFMDYFSEBFNRNDFLELLRRRTGTRKVHNNIVYNEYISHREIHMNATOWETLDTFK	120
Db	61	PQOFMDYFSEBFNRNDFLELLRRRTGTRKVHNNIVYNEYISHREIHMNATOWETLDTFK	120
Qy	121	WLREGGLCKVDETPKGYIYQYIDRDPETIRQLELEKKKKQDLDDEETAKFIEBQVRG	180
Db	121	WLREGGLCKVDETPKGYIYQYIDRDPETIRQLELEKKKKQDLDDEETAKFIEBQVRG	180
Qy	181	LEGKEQETPVFTELSRNEEBEKVFNLNKGAGGSAGATTSSSSLGPSALKILGSAASGK	240
Db	181	LEGKEQETPVFTELSRNEEBEKVFNLNKGAGGSAGATTSSSSLGPSALKILGSAASGK	240
Qy	241	RKSSQSSAPAKKKKSGALDIMELEBEEKKTARTDAPLQPGIVVKYITKKLGEKYHKK	300
Db	241	RKSSQSSAPAKKKKSGALDIMELEBEEKKTARTDAPLQPGIVVKYITKKLGEKYHKK	300
Qy	301	GVVKEVIDRYTAVVYKMTDSGDRKLQDTHLETVPAPGKRVLVNLNGGYRGNEGTLESINE	360
Db	301	GVVKEVIDRYTAVVYKMTDSGDRKLQDTHLETVPAPGKRVLVNLNGGYRGNEGTLESINE	360
Qy	361	KAFSATIVETGPKLGRRVEGIOYEDISKLA	391
Db	361	KAFSATIVETGPKLGRRVEGIOYEDISKLA	391

[illegible]

PA (BUCK-) BUCK INST AGE RES.
 XX Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;
 PI Warnock DE;
 XX WPI; 2003-845369/78.
 XX Identifying a mitochondrial target for drug screening assays and for
 PT treating diseases associated with altered mitochondrial function,
 PT comprises detecting a modified polypeptide in a sample and correlating
 PT with the disease.
 XX Claim 1; SEQ ID NO 1465; 180pp; English.
 XX This invention relates to novel mitochondrial targets that can be used
 CC for therapeutic intervention in treating a disease associated with
 CC altered mitochondrial function. Specifically, it refers to a method for
 CC identifying proteins of the human heart mitochondrial proteome that are
 CC useful for drug screening assays, as well as therapeutic targets. The
 CC present invention describes a method for identifying such proteins that
 CC can be used in the treatment of various diseases associated with altered
 CC mitochondrial function including diabetes mellitus, Huntington's disease,
 CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
 CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy
 CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
 CC compositions have neuroprotective, nootropic, antidiabetic,
 CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and
 CC cytoskeletal activities. This polypeptide sequence is a human heart
 CC mitochondrial protein of the invention.
 XX Sequence 393 AA;

Query Match 93.4%; Score 1895; DB 7; Length 393;
 Best Local Similarity 92.4%; Pred. No. 1.8e-159;
 Matches 363; Conservative 14; Mismatches 14; Indels 2; Gaps 1;

Qy 1 MGKSDFLSPKAIANRIKSGLOKLRWYCOMCQCRDENGFKCHCMSESHORQLLAS 60
 Db 1 MGKSDFLTPKAIANRIKSGLOKLRWYCOMCQCRDENGFKCHCMSESHORQLLAS 60
 Qy 61 PQQFMDYFSEFRNDFLELLRRFGTKRVHNNIVYNEYISHREIHNNATQWETLDTFK 120
 Db 61 PQQFMDYFSEFRNDFLELLRRFGTKRVHNNIVYNEYISHREIHNNATQWETLDTFK 120
 Qy 121 WLGRGLCKVDTPKGYIYQIDRDPETIRROLELEKKKKQDLDEEKTAKFIEQVRRG 180
 Db 121 WLGRGLCKVDTPKGYIYQIDRDPETIRROLELEKKKKQDLDEEKTAKFIEQVRRG 180
 Qy 181 LEGKEQTPVFTSLRENEBEKVTFNLKNGAGGSAGATTSSSLGPSALKLGSASGK 240
 Db 181 LEGKEQTPVFTSLRENEBEKVTFNLKNGAGGSAGATTSSSLGPSALKLGSASGK 240
 Qy 241 RKSSQSSAOPAKKKKSSALDEIMELEBEKKRTARTDAMLPQGIIVKIIITKKLGKHYHK 298
 Db 241 RKSSQSSQTSQSEKKKKSSALDEIMELEBEKKRTARTDAMLPQGIIVKIIITKKLGKHYHK 300
 Qy 299 KKGVKEVIDRYTAVVXMTDSGDRKLQDTHLETVPAPGRKRVVLNNGVYRGNEGTL 358
 Db 301 KKAIVKEVIDRYTAVVXMTDSGDRKLQDTHLETVPAPGRKRVVLNNGVYRGNEGTL 360
 Qy 359 NEKAFSATIVETGPKGRRVEGIQYEDISKLA 391
 Db 361 NEKTFSATIVETGPKGRRVEGIQYEDISKLA 393

RESULT 4
 AAY23617
 ID AAY23617 standard; protein; 291 AA.
 XX
 AC AAY23617;
 XX
 DT 07-SEP-1999 (first entry)
 XX

DE Mouse truncated kin17 protein.
 XX kin17 protein; cell proliferation; fertility; hyperproliferative disease;
 KW protein interaction; curved DNA; HIV replication; HIV integration;
 KW repair enzyme.
 XX Mus sp.
 OS
 PN FR2772046-A1.
 XX
 PD 11-JUN-1999.
 XX
 PF 09-DEC-1997; 97FR-00015536.
 XX
 PR 09-DEC-1997; 97FR-00015536.
 XX
 PA (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
 XX
 PI Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
 XX WPI; 1999-359999/31.
 DR
 PT New DNA coding for human kin17 protein - useful for controlling cell
 PT proliferation or fertility.
 XX
 PS Claim 14; Page 33-34; 69pp; French.
 XX
 CC The present sequence represents a mouse kin17 protein with amino acids
 CC 129-228 deleted. The mammalian kin17 protein is useful for preparing a
 CC medicament for controlling cell proliferation or for controlling
 CC fertility. The medicaments can also be used to treat hyperproliferative
 CC diseases. Fragments between amino acids 55 and 235 (preferably between
 CC amino acids 129 and 228) of a mammalian kin17 protein are useful for
 CC regulating the interaction between proteins and curved DNA. The fragment
 CC can be used to block replication of HIV or its integration into the human
 CC genome or to target repair enzymes to curved DNA sites. Expression
 CC vectors for kin17 can be used for controlling cell proliferation
 XX
 SQ Sequence 291 AA;

Query Match 71.8%; Score 1456; DB 2; Length 291;
 Best Local Similarity 74.4%; Pred. No. 1.3e-120;
 Matches 291; Conservative 0; Mismatches 0; Indels 100; Gaps 1;

Qy 1 MGKSDFLSPKAIANRIKSGLOKLRWYCOMCQCRDENGFKCHCMSESHORQLLAS 60
 Db 1 MGKSDFLSPKAIANRIKSGLOKLRWYCOMCQCRDENGFKCHCMSESHORQLLAS 60
 Qy 61 PQQFMDYFSEFRNDFLELLRRFGTKRVHNNIVYNEYISHREIHNNATQWETLDTFK 120
 Db 61 PQQFMDYFSEFRNDFLELLRRFGTKRVHNNIVYNEYISHREIHNNATQWETLDTFK 120
 Qy 121 WLGRGLCKVDTPKGYIYQIDRDPETIRROLELEKKKKQDLDEEKTAKFIEQVRRG 180
 Db 121 WLGRGLCKVDTPKGYIYQIDRDPETIRROLELEKKKKQDLDEEKTAKFIEQVRRG 180
 Qy 181 LEGKEQTPVFTSLRENEBEKVTFNLKNGAGGSAGATTSSSLGPSALKLGSASGK 240
 Db 129 -----ALKLGSAASGK 140
 Qy 241 RKSSQSSAOPAKKKKSSALDEIMELEBEKKRTARTDAMLPQGIIVKIIITKKLGKHYHK 300
 Db 141 RKSSQSSAOPAKKKKSSALDEIMELEBEKKRTARTDAMLPQGIIVKIIITKKLGKHYHK 200
 Qy 301 GVKKEVIDRYTAVVXMTDSGDRKLQDTHLETVPAPGRKRVVLNNGVYRGNEGTL 360
 Db 201 GVKKEVIDRYTAVVXMTDSGDRKLQDTHLETVPAPGRKRVVLNNGVYRGNEGTL 260
 Qy 361 KAFSATIVETGPKGRRVEGIQYEDISKLA 391
 Db 261 KAFSATIVETGPKGRRVEGIQYEDISKLA 291

RESULT 5
AAY23618
ID AAY23618 standard; protein; 293 AA.
XX
XX AAY23618;
AC
XX
DT 07-SEP-1999 (first entry)
XX
DE A huamn truncated kin17 protein.
XX
XX kin17 protein; cell proliferation; fertility; hyperproliferative disease;
KW protein interaction; curved DNA; HIV replication; HIV integration;
KW repair enzyme.
XX
OS Homo sapiens.
XX
PN FR2772046-A1.
XX
PD 11-JUN-1999.
XX
PF 09-DEC-1997; 97FR-00015536.
XX
PR 09-DEC-1997; 97FR-00015536.
XX
XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
XX
XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
PI WPI; 1999-359999/31.
XX
DR New DNA coding for human kin17 protein - useful for controlling cell
PT proliferation or fertility.
XX
XX Claim 15; Page 34-35; 69pp; French.
PS
CC The present sequence represents a truncated human kin17 protein with
CC amino acids 128-228 deleted. The mammalian kin17 protein is useful for
CC preparing a medicament for controlling cell proliferation or for
CC controlling fertility. The medicaments can also be used to treat
CC hyperproliferative diseases. Fragments between amino acids 55 and 235
CC (preferably between amino acids 129 and 228) of a mammalian kin17 protein
CC are useful for regulating the interaction between proteins and curved
CC DNA. The fragment can be used to block replication of HIV or its
CC integration into the human genome or to target repair enzymes to curved
CC DNA sites. Expression vectors for kin17 can be used for controlling cell
CC proliferation
XX
SQ Sequence 293 AA;

Query Match 67.3%; Score 1365; DB 2; Length 293;
Best Local Similarity 69.2%; Pred. No. 1.6e-112;
Matches 272; Conservative 9; Mismatches 10; Indels 102; Gaps 2;
Qy 1 MGKSDFLSPKATANRIKSKGLQKLRWYCMQCQCRDENGFKCHMSHQRQLLASN 60
Db 1 MGKSDFLTPKATANRIKSKGLQKLRWYCMQCQCRDENGFKCHMSHQRQLLASN 60
Qy 61 PQQFMDYFSEEFNDLFLRLRRFGTKRVHNNIVYNEYISHREHIHMNATQWETLTDFTK 120
Db 61 PQQFMDYFSEEFNDLFLRLRRFGTKRVHNNIVYNEYISHREHIHMNATQWETLTDFTK 120
Qy 121 WLGRGLCKVDYETPKGWYIQYIDRDPETTRRQLEKKKKQDLDDEKTAKEEQVRG 180
Db 121 WLGRGLC----- 128
Qy 181 LEGKEQETPVFTELSRNEEEKVTFTNLKAGGAGATTSSKSSSLGPSALKLLGSAAGK 240
Db 129 -----ALKTIGSASVK 140
Qy 241 RKSSOSSAQ--PAKKKKSALDSEIMEBEKKKRTARTDQWLGPIGVVVKIITKKLGSKYHK 298
Db 141 RKSSQSSSTQSKKKKKKALDSEIMEBEKKKRTARTDQWLGPIGVVVKIITKKLGSKYHK 200

Qy 299 KGVVKEVIDRYTAVVMTDSGDRLLKLDQTHLETVIPAGKRVLVNGYRGNEGTLSEI 358
Db 201 KKAIVKEVIDRYTAVVMTDSGDKLKDQTHLETVIPAGKRVLVNGYRGNEGTLSEI 260
Qy 359 NEKAFSATIVTETGPKGRRVEGIIQYEDISKLA 391
Db 261 NEKTFATIVTETGPKGRRVEGIIQYEDISKLA 293
RESULT 6
ADP22452
ID ADP22452 standard; protein; 390 AA.
XX
AC ADP22452;
XX
DT 12-AUG-2004 (first entry)
XX
DE Sea-squirt (Ciona intestinalis) zinc finger protein #13.
XX
KW sea-squirt; zinc finger protein; gene detection; drug development;
KW zinc finger protein-associated disease.
XX
OS Ciona intestinalis.
XX
PN JP2004057126-A.
XX
PD 26-FEB-2004.
XX
PF 31-JUL-2002; 2002JP-00222484.
XX
PR 31-JUL-2002; 2002JP-00222484.
XX
PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX
DR WPI; 2004-208711/20.
DR N-PSDB; ADP22451.
XX
PT Novel gene encoding zinc finger protein, useful as probe in gene
PT detecting instruments and in development of drug for treating zinc finger
PT protein associated diseases.
XX
PS Claim 1; SEQ ID NO 26; 972pp; Japanese.
XX
CC The invention comprises the amino acid and coding sequences of sea-squirt
CC (Ciona intestinalis) zinc finger proteins. The DNA and protein sequences
CC of the invention are useful in a gene detecting instrument. The DNA and
CC protein sequences of the invention are useful in the development of drugs
CC for the treatment of zinc finger protein-associated diseases. The present
CC amino acid sequence represents a sea-squirt zinc finger protein of the
CC invention.
XX
SQ Sequence 390 AA;

Query Match 61.4%; Score 1245; DB 8; Length 390;
Best Local Similarity 62.8%; Pred. No. 1.1e-101;
Matches 247; Conservative 50; Mismatches 84; Indels 12; Gaps 6;
Qy 1 MGKS--DFLSPKAIANRIKSKGLQKLRWYCMQCQCRDENGFKCHMSHQRQLLAS 58
Db 1 MGKEKPGELTAKAIGNRIKSKGLQKLRWYCMQCQCRDENGFKCHMSHQRQLLAS 60
Qy 59 ENPQFMDYFSEEFNDLFLRLRRFGTKRVHNNIVYNEYISHREHIHMNATQWETLTD 118
Db 61 ENPGAFHSFFSSFFDMQLLTKTFGKRVHNNIVYNEYISHREHIHMNATQWETLTD 120
Qy 119 TKWLREGGLCKVDYETPKGWYIQYIDRDPETTRRQLEKKKKQDLDDEKTAKEEQVR 178
Db 121 TKWLREGGHCKVDQTEKGFQYIDRDLVLARQELDKQKQREDDDERQAKAIEEMVK 180
Qy 179 RGLG-GRKEQETPVFTELSRNEEEKVTFTNLKAGGAGATTSSKSSSLGPSALKLLGSA 237
Db 181 RQDTAKDSGKFTLVLR-NDEEKIATNLQWNSPKA-----ATSSQLAP---RVSLSL 232

Db 61 PGKLFHSFKESFGSDGYWELLRRFGTKRTSANKIYQEVIAHKEHIHMNATRWLTLSLDYVK 121

Qy 121 WLGREGLCKVDETPKGYIOYIDRDPETIIRQLELEKKKODLDEKTKAFIEQVR-- 178

Db 121 WLGRGTQVIADETEGKGFVTVIDRSPAMERQAKADRKEMKDEDERMADFIEQIQNA 180

Qy 179 RLGEGKEOE--TPVFTELSR--ENEEEKVTFLNKGAGGSAGATTSSKSSSLGPSALKLGS 236

Db 181 KAKDGEDEGEQKFTELKREENBPLKDJRILEK-----KFQPDVTVLGKSALAKRPAP 232

Qy 237 ASGKRKRESSQSAQPAKKKKKALSALDEIMELBEEKK-RTARTDAMLQPGIVVKKIITKKLGEK 295

Db 233 EAEKVFKKPKSVAGDSQTSVLDEIYKQESKKERANKDYWLHKGIWVKFISKSMGEK 292

Qy 296 YHKKKGVVKEVIDRYTAVKWTSGDBLKLDQTHLETVIPAGKRVVLVNGGYRGNEGTL 355

Db 293 FFKQKAVLVVIDRYQGKIRFLETGEEKLVDAQHLETVIPALDKPVMVYVNGAYRGSEALL 352

Qy 356 ESINEKAFSATIVETGTPKGRVVEGIOYEDISKL 390

Db 353 RKLDERYSVSVEILHGLPLKGRIVDNVQYEDISKL 387

RESULT 8

AAG42576

ID AAG42576 standard; protein; 411 AA.

XX AAG42576;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 53114.

DE Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

XX Arabidopsis thaliana.

XX EPI033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.

PR 25-MAR-1999; 99US-0126284P.

PR 29-MAR-1999; 99US-0126785P.

PR 01-APR-1999; 99US-0127462P.

PR 06-APR-1999; 99US-0128234P.

PR 08-APR-1999; 99US-0128714P.

PR 16-APR-1999; 99US-0129845P.

PR 19-APR-1999; 99US-0130077P.

PR 21-APR-1999; 99US-0130449P.

PR 23-APR-1999; 99US-0130510P.

PR 23-APR-1999; 99US-0130891P.

PR 28-APR-1999; 99US-0131449P.

PR 30-APR-1999; 99US-0132048P.

PR 30-APR-1999; 99US-0132407P.

PR 04-MAY-1999; 99US-0132484P.

PR 05-MAY-1999; 99US-0132485P.

PR 06-MAY-1999; 99US-0132486P.

PR 06-MAY-1999; 99US-0132487P.

PR 07-MAY-1999; 99US-0132863P.

PR 11-MAY-1999; 99US-0134256P.

PR 14-MAY-1999; 99US-0134218P.

PR 14-MAY-1999; 99US-0134219P.

PR 14-MAY-1999; 99US-0134221P.

PR 14-MAY-1999; 99US-0134370P.

PR 18-MAY-1999; 99US-0134768P.

PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142220P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.

PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-015753P.
PR 07-OCT-1999; 99US-0157865P.
PR 08-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.

```

PR 26-OCT-1999; 99US-01611360P.
PR 26-OCT-1999; 99US-01611361P.
PR 28-OCT-1999; 99US-01611920P.
PR 28-OCT-1999; 99US-01611922P.
PR 28-OCT-1999; 99US-01611933P.
PR 28-OCT-1999; 99US-0162142P.

Query Match 48.7%; Score 988; DB 3; Length 411;
Best Local Similarity 48.8%; Pred. No. 8.7e-79;
Matches 203; Conservative 67; Mismatches 116; Indels 30; Gaps 7;

QY 1 MGKSDFLSPKAIANRIKSKGLQKLRWYQCMQCKQCRDENGFKCHCMSESHORQLLASEN 60
DB 1 MGKNDFLTPTKAIANRIKAKGLQKLRWYQCMQCKQCRDENGFKCHCMSESHORQMQVFGQN 60

QY 61 PQQFMDYFSEFRNDLFLRLRRFGTKRVNNIVYNEYISHREHNNATQWETLTDFTK 120
DB 61 PTRVVDGYSEFEQTFDLMLRRSHRFSRIAATVVYNEYINDRHVHNNSTEWATLTFEIK 120

QY 121 WLGRGLCKVDETPKGYIOYIDRDPETIRROLELEKKKODLDEKTAKEEQVRRG 180
DB 121 HLKGTGCKVEETPKGYIYIDRDESETLFRKLNKRVKSDLAEBEKQREBQRIQIERA 180

QY 181 LE-----GKEQETPVFTLSRENEE-----KVTNLNKGAGGSAGATTSKS 222
DB 181 AEKLNCGGGEGETSGNDEVDDGDERKDKEDLRLKSGVKVGFAL---GGGVKQVATGKE 237

QY 223 SSLGPSALKULGSAAS-----GKRKSSQSSAQAPAKKKKALDEIMELEBEKK-RTARTD 276
DB 238 R--GESSKLLFGDBENDKVERGEKRRKSGDGRSEKERRSALDELAKKEBEKKERMNRKD 295

QY 277 AWLQPGIVVKIITKKLGEK-YHKKGKGVVKEVIDRYTAVVKTQSGDRLKLDQTHLETVIP 335
DB 296 YWLFEGHIVKMSKALAEKGYIKQGVVKKVIDNYGIEIKMLDSKHVLRVDQKELETVLP 355

QY 336 APGRKRVLVNGYRGNEGTLSEINEKAFSATIVTETGLKGRRVEGIQYEDISKLA 391
DB 356 QIGGMVKIVNGAYRGSNARLLGVDTKEFCAKVQIEKGYDGRVTKSIEYEDICKLA 411

RESULT 9
AAB03063
ID AAB03063 standard; protein; 424 AA.
XX
AC AAB03063;
XX
DT 27-SEP-2000 (first entry)
XX
DE Maize KIN17 orthologue, ZmKINH-1.
XX
KW ZmKINH-1; KIN17 orthologue; maize; zinc finger protein; RecA homologue;
KW nuclear localisation; nonhomologous recombination;
KW illegitimate recombination; double stranded DNA binding; curved DNA;
KW homologous gene targeting; transgenic plant.
XX
OS Zea mays.
XX
PN WO200024900-A1.
XX
PD 04-MAY-2000.
XX
PF 06-OCT-1999; 99WO-US023280.
XX
PR 27-OCT-1998; 98US-0105802P.
XX
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Mahajan PB;
XX
WPI; 2000-350741/30.
DR N-PSDB; AAA52589.
XX
Nucleic acids encoding maize KIN17 orthologue proteins useful for

```

```

PT preventing illegitimate recombination in cells.
XX Claim 11; Page 64-66; 84pp; English.
XX
CC This sequence represents the maize KIN17 orthologue ZmKINH-1. The
CC invention relates to maize KIN17 orthologues (AAB03063-B03065), nucleic
CC acids encoding them (AAA52589-A52591), and expression vectors, transgenic
CC plants and plant seeds comprising nucleotides encoding maize KIN17
CC orthologues. KIN17 has, until now, been found only in animal (avian,
CC rodent and human) cells, this invention being the first report describing
CC the presence of KIN17 in plants. Murine KIN17 was found to have
CC significant homology to Escherichia coli RecA protein, and contains a
CC zinc finger motif and a nuclear localisation signal. KIN17 binds double-
CC stranded DNA, preferentially binding to curved DNA, and forms
CC intranuclear foci on overexpression in mammalian cells. It is also
CC indicated that KIN17 plays a role in non-homologous (illegitimate)
CC recombination, which occurs at higher rates among higher eukaryotes,
CC particularly plants. Illegitimate recombination in plants is a major
CC impediment to the generation of transgenic crops such as maize. Maize
CC KIN17 orthologue nucleic acid sequences may be used to generate
CC transgenic plants. The transgenic plants generated can be monocots or
CC dicots and are particularly maize, soybean, sunflower, sorghum, canola,
CC wheat, alfalfa, cotton, rice, barley and millet. In particular, antisense
CC sequences may be used to reduce KIN17 levels in embryogenic callus or
CC embryo cells, thereby reducing the amount of non-homologous recombination
CC and enhancing homologous gene targeting
XX
SQ Sequence 424 AA;

Query Match 45.0%; Score 913; DB 3; Length 424;
Best Local Similarity 44.9%; Pred. No. 4.3e-72;
Matches 190; Conservative 75; Mismatches 126; Indels 32; Gaps 7;

QY 1 MGKSDFLSPKAIANRIKSKGLQKLRWYQCMQCKQCRDENGFKCHCMSESHORQLLASEN 60
DB 1 MGKHEFLTPTKAIANRIKAKGLQKLRWYQCMQCKQCRDENGFKCHCMSESHORQMQVFGMA 60

QY 61 PQQFMDYFSEFRNDLFLRLRRFGTKRVNNIVYNEYISHREHNNATQWETLTDFTK 120
DB 61 PDRVVEGFSEEFLESFLIRRAHRSRVAAVTVYNEYIADRHVHNNSTFWATLTFFVK 120

QY 121 WLGRGLCKVDETPKGYIOYIDRDPETIRROLELEKKKODLDEKTAKEEQVRR- 179
DB 121 FLGREGYCKVEDTPKGYIYIDRDESETLFRKLNKRVKSDLAEBEKQREBQRIQIERA 180

QY 180 -----GLEGKEQETPVFTLSRENEEKTFFNLNKGAGGSAG--ATTSSSLGSP 227
DB 181 HKSLAKPNGGGAAGEPESEGESGDDDEPEDDSKEADKATKIALQKAVPGP 240

QY 228 SA-----LKL-----LGSASGKRK---SSSQSSAQAPAKKKKALDEIMELEBE-K 269
DB 241 KVPFDDPKPKGFDDEDSGTRDQKNELTKMGKQKDVAAEAKRSALDELMEKEBEKAK 300

QY 270 KRTARTDAWLQPGIVVKIITKKLGEK-YHKKGKGVVKEVIDRYTAVVKTQSGDRLKLDQ 328
DB 301 ERSNRKDYWLCPGIVVKVMSKLAEGYIKQGVVKKVIDNYGIEIKMLDSKHVLRVDQ 360

QY 329 HLETVIPAPGRVVLVNGYRGNEGTLSEINEKAFSATIVTETGLKGRRVEGIQYEDIS 388
DB 361 ELETVIPQIGLVRIVNGAYRGSNARLLSDVTEKFCAKVQVEKGLYDGKVLRAVEYEDIC 420

QY 389 KLA 391
DB 421 KIS 423

RESULT 10
ADX96318
ID ADX96318 standard; protein; 437 AA.
XX
AC ADX96318;
XX

```


PS Claim 1; SEQ ID NO 48183; 15pp; English.

XX The invention describes a recombinant DNA construct comprising a

CC polynucleotide consisting of a sequence encoding an amino acid sequence

CC available in electronic form from the US patent office at

CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide

CC of the invention are also useful in physical arrays of molecules and as

CC plant breeding markers. The recombinant DNA construct is useful for

CC improving plant tolerance to cold, heat, drought, herbicides, extreme

CC osmotic conditions, pathogens or pests, for manipulating growth rate in

CC plant cells by modification of the cell cycle pathway, for conferring

CC increased resistance to plant disease, for producing galactomannan,

CC lignin or plant growth regulators, for increasing the rate of homologous

CC recombination in plants, for improving yield by modification of

CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake

CC or by providing improved plant growth and development under at least one

CC stress condition or for modifying seed oil or protein yield and/or

CC content. This is the amino acid sequence of a plant full length insert

CC polypeptide that can be used in the recombinant DNA construct of the

XX invention.

SQ Sequence 437 AA;

Query Match 45.0%; Score 913; DB 8; Length 437;

Best Local Similarity 44.9%; Pred. No. 4.5e-72;

Matches 190; Conservative 75; Mismatches 126; Indels 32; Gaps 7;

Qy 1 MGKSDFLSPKAIANRIKSKGLQKLRYWCQKQKCRDENGPKCHCMSESHORQLLASEN 60

Db 14 MGKHEFTPKAIANRIKAKGLQKLRYWCQKQKCRDENGPKCHCMSESHORQMVFGMA 73

Qy 61 PQQFMDVFSFERNDFLELLRRRFGTKRVHNNIYVYNEIISHREIHNATOWETLTDFTK 120

Db 74 PDRVVEGSEEFLESLIRRAHRHGRVAATVYVYNEIADRHVHNNSTRWATLTDFVK 133

Qy 121 WLREGKCKVDTPKGVYQIYDIPETIRROLEKLEKKKQDLDDEKTKAFIEQVRR- 179

Db 134 FLRGEGYCKVEDTPKGVFWTYIDRSEQAVKRLKRIKSDMADDERQERMIARQIERA 193

Qy 180 -----GLEGEQETPVFTLSRENEBEKVTNFKAGGSAG--ATTSSSSSLGP 227

Db 194 HKSLAKPNGGAAEGEPESGSEYSGSDDEQBPEDDSKEADKATGKATGATATLQKAVRGP 253

Qy 228 SA-----LKL-----LGAASGKRK-----SSQSAQPAKKKKSALDIMELEE-K 269

Db 254 KVNPFDDPKPKFGFDEEDSGTRDQEKNELTQKMGKDKVKAARSAALDELMEKEERAK 313

Qy 270 KRTARTDALQPGIIVKIIITKLGK-YHKKGVVKEVIDRYTAVVKMTDSGDRLLKLDQT 328

Db 314 ERSNRKDYWLCPGIVVKVNSKSLAEKGYKQKGVVKKVIDKYVGEIEMLESHVLRVDQD 373

Qy 329 HLETVIPAGKRVLVNGYRGNEGTLSINEKAFSATIVETGPLKRRVVEGIQYEDIS 388

Db 374 ELETVIPQIGLVYNGAYRGSNARLLSVDETEKCAKVQVEKGLYDGKVLRAVEYEDIC 433

Qy 389 KLA 391

Db 434 KIS 436

RESULT 12

AAB03064

ID AAB03064 standard; protein; 423 AA.

XX AAB03064;

AC AAB03064;

XX 27-SEP-2000 (first entry)

DE Maize KIN17 orthologue, ZmKINH-2.

KW ZmKINH-2; KIN17 orthologue; maize; zinc finger protein; RecA homologue;

XX nuclear localisation; nonhomologous recombination;

KW illegitimate recombination; double stranded DNA binding; curved DNA;

homologous gene targeting; transgenic plant.

Zea mays.

WO2000024900-A1.

04-MAY-2000.

06-OCT-1999; 99WO-US023280.

27-OCT-1998; 98US-0105802P.

(PION-) PIONEER HI-BRED INT INC.

Mahajan PB;

WPI; 2000-350741/30.

N-PSDB; AAB52590.

Nucleic acids encoding maize KIN17 orthologue proteins useful for preventing illegitimate recombination in cells.

Claim 11; Page 64-66; 84pp; English.

This sequence represents the maize KIN17 orthologue ZmKINH-2. The invention relates to maize KIN17 orthologues (AAB03063-B03065), nucleic acids encoding them (AAB52589-A52591), and expression vectors, transgenic plants and plant seeds comprising nucleotides encoding maize KIN17 orthologues. KIN17 has, until now, been found only in animal (avian, rodent and human) cells, this invention being the first report describing the presence of KIN17 in plants. Murine KIN17 was found to have a significant homology to Escherichia coli RecA protein, and contains a zinc finger motif and a nuclear localisation signal. KIN17 binds double-stranded DNA, preferentially binding to curved DNA, and forms intranuclear foci on overexpression in mammalian cells. It is also induced on exposure to gamma or ultraviolet radiation. These findings indicate that KIN17 plays a role in non-homologous (illegitimate) recombination, which occurs at higher rates among higher eukaryotes, particularly plants. Illegitimate recombination in plants is a major impediment to the generation of transgenic crops such as maize. Maize KIN17 orthologue nucleic acid sequences may be used to generate transgenic plants. The transgenic plants generated can be monocots or dicots and are particularly maize, soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton, rice, barley and millet. In particular, antisense sequences may be used to reduce KIN17 levels in embryogenic callus or embryo cells, thereby reducing the amount of non-homologous recombination and enhancing homologous gene targeting

Sequence 423 AA;

Query Match 44.5%; Score 901.5; DB 3; Length 423;

Best Local Similarity 44.5%; Pred. No. 4.5e-71;

Matches 189; Conservative 72; Mismatches 127; Indels 37; Gaps 8;

Qy 1 MGKSDFLSPKAIANRIKSKGLQKLRYWCQKQKCRDENGPKCHCMSESHORQLLASEN 60

Db 1 MGKHEFTPKAIANRIKAKGLQKLRYWCQKQKCRDENGPKCHCMSESHORQMVFGMA 60

Qy 61 PQQFMDVFSFERNDFLELLRRRFGTKRVHNNIYVYNEIISHREIHNATOWETLTDFTK 120

Db 61 PDRVVEGSEEFLESLIRRAHRHGRVAATVYVYNEIADRHVHNNSTRWATLTDFVK 120

Qy 121 WLREGKCKVDTPKGVYQIYDIPETIRROLEKLEKKKQDLDDEKTKAFIEQVRRG 180

Db 121 LLRGEGYCKVEDTPKGVFWTYIDRSEQAVKRLKRIKSDMADDERQERMIARQIERA 180

Qy 181 LE-----GKEQETPVFTLSRENEE-EKVTNFKAGGSAG--TTS 220

Db 181 HKSLAKPNGGAAEGEPESGSEYSGSDDEQBPEDDSKEADKATGKATGATATLQKAV 237

Qy 221 KSSSLGSPALK-----LGAASGKRK-----SSQSAQPAKKKKSALDIMELEE- 267

Db 238 RGPKNPPEKPEVKRFRFDEEDLGTGEEKDELAKKKGDKVKAARSAALDELMEKEER 297

XX New recombinant DNA constructs useful in the field of biochemistry and
PT genetics, and in particular for producing transgenic plants with improved
PT biological characteristics.
PT

Db	298	AKERSNRKDYMLCPGIVVVKVMSKSLAEKGYKQKGVVRKVMKDYVGEIEMLESKHVLRVD	35
Qy	327	QTHLETVIPAGKRVVLNGYRGNEGTLSEINEKAFSATIVLETGPKGRRVEGIQYED	386
Db	358	QDELSTVLPQTGGVRLNGAYRGSNARLLSVDTKEFCAKVQVEKGLYDGKVLRAVKYED	417
Qy	387	ISKLA 391	
Db	418	ICKIS 422	
RESULT 15			
AAG42577			
ID	AAG42577 standard; protein; 382 AA.		
XX	AAG42577;		
DT	18-OCT-2000 (first entry)		
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 53115.		
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 53115.		
XX	Protein identification; signal transduction pathway; metabolic pathway;		
KW	hybridisation assay; genetic mapping; gene expression control; promoter;		
KW	termination sequence.		
XX	XX		
OS	Arabidopsis thaliana.		
XX	XX		
PN	EP1033405-A2.		
XX	XX		
PD	06-SEP-2000.		
XX	XX		
PF	25-FEB-2000; 2000EP-00301439.		
XX	XX		
PR	25-FEB-1999; 99US-0121825P.		
PR	05-MAR-1999; 99US-0123180P.		
PR	09-MAR-1999; 99US-0123548P.		
PR	23-MAR-1999; 99US-0125788P.		
PR	23-MAR-1999; 99US-0126264P.		
PR	29-MAR-1999; 99US-0126785P.		
PR	01-APR-1999; 99US-0127462P.		
PR	06-APR-1999; 99US-0128234P.		
PR	08-APR-1999; 99US-0128714P.		
PR	16-APR-1999; 99US-0129845P.		
PR	13-APR-1999; 99US-0130077P.		
PR	21-APR-1999; 99US-0130443P.		
PR	23-APR-1999; 99US-0130510P.		
PR	23-APR-1999; 99US-0130891P.		
PR	28-APR-1999; 99US-0131449P.		
PR	30-APR-1999; 99US-0132048P.		
PR	30-APR-1999; 99US-0132407P.		
PR	04-MAY-1999; 99US-0132484P.		
PR	05-MAY-1999; 99US-0132485P.		
PR	06-MAY-1999; 99US-0132486P.		
PR	07-MAY-1999; 99US-0132487P.		
PR	11-MAY-1999; 99US-0132863P.		
PR	14-MAY-1999; 99US-0134256P.		
PR	14-MAY-1999; 99US-0134218P.		
PR	14-MAY-1999; 99US-0134219P.		
PR	14-MAY-1999; 99US-0134221P.		
PR	18-MAY-1999; 99US-0134370P.		
PR	19-MAY-1999; 99US-0134768P.		
PR	20-MAY-1999; 99US-0134941P.		
PR	20-MAY-1999; 99US-0135124P.		
PR	21-MAY-1999; 99US-0135353P.		
PR	24-MAY-1999; 99US-0135629P.		
PR	25-MAY-1999; 99US-0136021P.		
PR	27-MAY-1999; 99US-0136392P.		
PR	28-MAY-1999; 99US-0136782P.		
PR	01-JUN-1999; 99US-0137222P.		
PR	03-JUN-1999; 99US-0137528P.		
PR	04-JUN-1999; 99US-0137502P.		
PR	07-JUN-1999; 99US-0137724P.		
PR	08-JUN-1999; 99US-0138094P.		

PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 28-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142820P.
PR 12-JUL-1999; 99US-0142877P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 08-AUG-1999; 99US-0147493P.

PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-015753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158234P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 41.6%; Score 843; DB 3; Length 382;
Best Local Similarity 45.7%; Pred. No. 6.2e-66;
Matches 177; Conservative 64; Mismatches 116; Indels 30; Gaps 7;

Db	1	MCQKQCEDENGFKCHCNSHSHQMQVFGQNPTRVVDGYSEEFQTFDLMRSHRFSRI	60
Qy	90	HNNIVNEYISHREHIHNNATOWETLTDFTKWLGRBGLCKVDETPKGWYIOYIDRDPETI	149
Db	61	AATVVVNEYINDRRHHVHMNSTEWATLTFETIKHLGKTGCKVEETPKGWFITYIDRSETL	120
Qy	150	RRQLELEKKKQDLDDDEKTAKEEEOVRRLGLE-----GKEQETPVPTFELSRENEEE---	201
Db	121	FKERLKNKRVKSDLAEBEEKEREIQOIERAAEKLNGGGGEGETSGNDEVDDGDDERKK	180
Qy	202	-----KVTFLNKGAGGSAGATTSSKSSSLGPSALKLGSAA-----GKRRESSQ	246
Db	181	DEDLRLKSGVKVGFAL---GGVKQVATGKER--GESSKLLFGDEENDKVERGEKRRSG	235
Qy	247	SSAQPAKKKXSALDEIMELBEEK-RTARTDAMLQPGIVVKIITKYLGEK-YHKKKGWVK	304
Db	236	DSGRSEKERRSALDELMKEBEEKKERNRKYMLFEGIIVKVMSKALAEKGYKQKGVVK	295
Qy	305	EVIDRYTAVVKMTDSGDRXLDOETHLETVIPAGKRVLVNLGGYRGNEGTLSEINEKAFS	364
Db	296	KVIDNYVGEIKMLDSKHVLRVDQKELETVLPQIGGMVKIVNGAYRGSNARLLGVDTBKFC	355
Qy	365	ATIVIETGPLKGRRVEGIQYEDISKLA	391
Db	356	AKVQIEKGVYDGRVKSIEYEDICKLA	382

Search completed: November 23, 2005, 16:28:20
Job time : 134.167 secs

This Page Blank (usp1c)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:19:43 ; Search time 152.61 Seconds
(without alignments)
1807.628 Million cell updates/sec

Title: US-09-555-529-25

Perfect score: 2028

Sequence: 1 MGKSDPLSPKAIANRIKSG.....GPLKGRVRVGIQVEDISKLA 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2028	100.0	391	2	Q8K339_MOUSE	Q8K339 mus musculus
2	1895	93.4	393	2	Q60870_HUMAN	Q60870 homo sapien
3	1678	82.7	389	2	Q6GJ97_XENTR	Q6GJ97 xenopus tro
4	1646	81.2	387	2	Q68F56_XENLA	Q68F56 xenopus lae
5	1552	76.5	383	2	Q5RH25_BRARE	Q5RH25 brachydanio
6	1112	54.8	387	2	Q7PGA1_ANOGA	Q7PGA1 anopheles g
7	1056	52.1	195	2	Q9CV58_MOUSE	Q9CV58 mus musculus
8	1052.5	51.9	390	2	Q9VPH4_DROME	Q9VPH4 drosophila
9	1050.5	51.8	390	2	Q8SXK2_DROME	Q8SXK2 drosophila
10	988	48.7	411	2	Q9ZVU5_ARATH	Q9ZVU5 arabidopsis
11	969.5	47.8	404	2	Q9XWF2_CABEL	Q9XWF2 caenorhabdi
12	932.5	46.0	269	2	Q4TDV9_TETNG	Q4TDV9 tetraodon n
13	913	45.0	430	2	Q751U5_ORYSA	Q751U5 oryza sativ
14	839	41.4	402	2	Q9A9B7_ORYSA	Q9A9B7 oryza sativ
15	835	41.2	397	2	Q60L39_CAEBR	Q60L39 caenorhabdi
16	831.5	41.0	427	2	Q4S5G2_TETNG	Q4S5G2 tetraodon n
17	825	40.7	445	2	Q55D16_DICDI	Q55D16 dictyosteli
18	689.5	34.0	244	2	Q76926_DROME	Q76926 drosophila
19	649	32.0	178	2	Q6X190_DROYA	Q6X190 drosophila
20	648	32.0	378	2	Q4N9K7_THEPA	Q4N9K7 theileria p
21	645	31.8	441	2	Q4XDV6_PLACH	Q4XDV6 plasmodium
22	644	31.8	412	2	Q4UG63_THEAN	Q4UG63 theileria a
23	644	31.8	442	2	Q8IKG2_PLAF7	Q8IKG2 plasmodium
24	644	31.8	445	2	Q4Z7D7_PLABE	Q4Z7D7 plasmodium
25	630	31.1	457	2	Q7RD49_PLAYO	Q7RD49 plasmodium
26	604	29.8	334	2	Q4WFG0_ASPFU	Q4WFG0 aspergillus
27	604	29.8	337	2	Q415W6_GIBZE	Q415W6 gibberella
28	590	29.1	335	2	Q5AQ80_EMENI	Q5AQ80 aspergillus
29	578.5	28.5	303	2	Q9Y7X9_SCHPO	Q9Y7X9 schizosacch
30	544	26.8	324	2	Q55M94_CRYNE	Q55M94 cryptococcu
31	544	26.8	324	2	Q5K8H7_CRYNE	Q5K8H7 cryptococcu

Q525W2 magnaporthe
Q4PG11 ustilago ma
Q6C2K1 yarrowia li
Q78742 neurospora
Q5CFM3 cryptospori
Q5CYD1 cryptospori
Q51F62 entamoeba h
Q6BQ00 debaryomyce
Q6CXJ7 kluyveromyce
Q759P8 ashbya goss
Q4XND8 plasmodium
Q5A673 candida alb
Q6FMS8 candida gla
Q4TDW0 tetraodon n

ALIGNMENTS

RESULT 1

Q8K339_MOUSE PRELIMINARY; PRT; 391 AA.

AC Q8K339; DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Antigenic determinant of rec-A protein.
GN Name=Kin;

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M.J., Udwin T.B., Toshiyuki S., Carninci P., Scheetz T.E.,
RA Brownstein M.J., Soares M.B., Bonaldo M.F., Casavant T.L., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

NUCLEOTIDE SEQUENCE.

RP STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

NUCLEOTIDE SEQUENCE.

RP MEDLINE=92020193; PubMed=1923796;
RX Angulo J., Rouer E., Mazin A., Mattei M., Tissier A., Horellou P.,
RA Benatous R., Devoret R.;
RT "Identification and expression of the cDNA of KIN17, a zinc-finger
gene located on mouse chromosome 2, encoding a new DNA binding
protein";
RL Nucleic Acids Res. 19:5117-5123(1991).
DR EMBL; BC028860; AAH28860.1; -; mRNA.
DR PIR; S18666; S18666.

Db 121 LGNGCHVADTDKGFYIYDRDPETLAMOQKWKQKMDKDAERLAEFFIEQVRRGK 180
 Qy 182 EGKEQBTVPFTELGRNEEKVTFNLKNGAGGAGATTSSKSLGPSAL----KLGSAS 238
 Db 181 TEEBPTCTSGVSELKRENEEDTIKELKLG-----SKQOQSTPSAVISKEPFDALDD 231
 Qy 239 GKREKSSQSAQAKKK-SALDIMELEEEK-RTARTDAWLOPGVVKVVKIKLGEKY 296
 Db 232 GKKEKKIKATSNGETKKGSALDELQEEQKKEKNRKYWLAEGLVVKLISRSLOEKY 291
 Qy 297 HKKGGVVKVEIDRYTAVVKNMDSGDRCLKDQTHLETIVPAGKRVLVNLNGYRGNEGTL 356
 Db 292 YKEGVVVEIEKYRAKIKLLETGERKLKVDQALHEIVIPAGQILVLNGYRGCTAVLK 351
 Qy 357 SINEKAFSATIVTETGLKRRVEGQYEDISK 390
 Db 352 AINTERYSVTETASGLKGLVSNVAYEDISK 385

RESULT 7

Q9CV58 MOUSE
 ID Q9CV58 MOUSE PRELIMINARY; PRT; 195 AA.
 AC Q9CV58;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DE Mus musculus adult male tongue cDNA, RIKEN full-length enriched
 DE library, clone:2310020017 product:antigenic determinant of rec-A
 DE protein, full insert sequence. (Fragment).
 GN Name=Kin;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Tongue;
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44(1999).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Tongue;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Aizawa K., Izawa M., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staebli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaudo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Guetincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata K., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [3]

NUCLEOTIDE SEQUENCE.

STRAIN=C57BL/6J; TISSUE=Tongue;

the FANTOM Consortium,

the RIKEN Genome Exploration Research Group Phase I & II Team;

"Analysis of the mouse transcriptome based on functional annotation of

60,770 full-length cDNAs.";
 Nature 420:563-573(2002).
 [4]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Tongue;
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1617-1630(2000).
 [5]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Tongue;
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Konno H., Akiyama J., Nishi K., Kiteunai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771(2000).
 [6]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Tongue;
 RX Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
 RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
 RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
 RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
 RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
 RA Okazaki Y., Okado T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
 RA Sogabe Y., Suzuki H., Tagami M., Tanaka A., Takahashi F., Tanaka T.,
 RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
 RA Muramatsu M., Hayashizaki Y.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK009429; BAB26281.2; -, mRNA.
 DR MGI; MGI:96676; Kin.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR InterPro; IPR007087; Znf_C2H2.
 DR SMART; SM00355; Znf_C2H2; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
 FT NON TER 195 195
 SQ SEQUENCE 195 AA; 23518 MW; 78112E9CA820BFA6 CRC64;

Query Match 52.1%; Score 1056; DB 2; Length 195;
 Best Local Similarity 100.0%; Pred. No. 2.5e-55;
 Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGKSDFLSPKAIANRIKSLGKLRLWYQCMQKQCRDENGFKCHCMSESHORQLLAS 60
 Db 1 MGKSDFLSPKAIANRIKSLGKLRLWYQCMQKQCRDENGFKCHCMSESHORQLLAS 60
 Qy 61 PQQPMDFSEBEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREIHNNATQWETLTDP 120
 Db 61 PQQPMDFSEBEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREIHNNATQWETLTDP 120
 Qy 121 WLGRGELCKVDETPKGYIQVDRDPETIRRLQLEKKKKQDLDEEKTAKFIEQVRG 180
 Db 121 WLGRGELCKVDETPKGYIQVDRDPETIRRLQLEKKKKQDLDEEKTAKFIEQVRG 180
 Qy 181 LEGKEQETPVFTELS 195
 Db 181 LEGKEQETPVFTELS 195

RESULT 8

```

O9VPH4 DROME
ID O9VPH4_DROME PRELIMINARY; PRT; 390 AA.
AC O9VPH4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2004 (TrEMBLrel. 26, Last annotation update)
DE CG5649+PA.
GN Name=kin17; ORFNames=CG5649;
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.D.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burris K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hostin D., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Llang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitzkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RN NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426065; PubMed=12537568;
RA Celnikier S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svitzkas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila
RT melanogaster euchromatic genome sequence."
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RN NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J.W., Svitzkas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celnikier S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:

```

```

a genomics perspective."
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).
RN [4]
RN NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celnikier S.E., de Grey A.D.N.-J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review."
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN [5]
RN NUCLEOTIDE SEQUENCE.
RG Berkeley Drosophila Genome Project;
RA Celnikier S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,
RA Hoskins R., Stapleton M., Pacleb J., Park S., Svitzkas R., Smith E.,
RA Yu C., Rubin G.;
RT "Drosophila melanogaster release 4 sequence."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RN NUCLEOTIDE SEQUENCE.
RG FlyBase;
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
CC -!- INTERACTION:
CC Q9VW16:CG1041; NbExp=1; IntAct=EBI-193999, EBI-175006;
CC Q9VPS5:HSP60B; NbExp=1; IntAct=EBI-193999, EBI-159512;
CC P21187:pAbp; NbExp=1; IntAct=EBI-193999, EBI-103658;
CC EMBL; AE003591; AAF51578.1; -, Genomic_DNA.
DR IntAct; Q9VPH4; -.
DR Ensembl; CG5649; Drosophila melanogaster.
DR FlyBase; FBgn0024887; CG5649.
DR FlyBase; FBgn0024887; kin17.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; P:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR007087; ZnF_C2H2.
DR SMART; SM00355; ZnF_C2H2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
KW Nuclear protein.
SQ SEQUENCE 390 AA; 45355 MW; 5F8231706C2D431F CRC64;
Query Match 51.9%; Score 1052.5; DB 2; Length 390;
Best Local Similarity 52.2%; Pred. No. 9.4e-55;
Matches 206; Conservative 69; Mismatches 107; Indels 13; Gaps 5;
QY 1 MKSDFLPKATANRIKSKGLQKRWYCOMCQKCRDENGFKCHCMSSHORQLLASN 60
DB 1 MGRAEVGTPKYLANKWKSKGLQKRWYCOMCEKCRDENGFKCHTWSHQLLAFDN 60
QY 61 PQOFMDYFSEEPNDFLELLRRRFGTKRVHNNIVVNYEISHREHIMNATOWETLTDFTK 120
DB 61 PGKFLHSFKSEFSDGCMELLRRRFGTKTSANKIYQEVIAHKEHIMNATRWLTLSDYK 120
QY 121 WLGREGLCVDETPKGVYQYIDRPETIRROLELEKKKKQDLDDEETAKTIEQVR-- 178
DB 121 WLGRTGQVIADETEGKGFVYIDRSPFAMERQAKADRKEMKEDDERMADIEQIKNA 180
QY 179 RLEGEKEQE-TPVFTFELSR-ENESEKVTFLNKNKGAGSGAGTTSKSSSGLPSALKLLGSA 236
DB 181 KAKDGEDEGGQKFTEFLKRENEPLKLDIRLEK-----KFQDPTVLGKSAKLRAP 232
QY 237 ASGKKESSQSSAQPAKKKKSALDEIMELEEK-K-RTARTDAWLQPGIIVVKITKKLGK 295
DB 233 EAEKVFKKPKSVAGDSQTSRSLVDRIIKQESKKEKRNKDYWLHKGIIVVKISKMGK 292
QY 296 YHKKGKGVKVEDRVTAVVQMTDSGRLKLDOTHLETVTPAPKRVVLVNGGYRNEGTL 355
DB 293 PFKQKAVLDVIDRYQGKIKFLETGKLVQDAHLETVTPALDVKPVVNVNGYRGSEALL 352

```



```

Qy 356 ESINEKAFSATVIETGPKLGRVVEGIQYEDISKL 390
Db 353 RKLDERYSVSVEILHGLPLKGRIVDNVQYEDISKL 387

RESULT 9
Q8SR2 DROME
ID Q8SR2 DROME PRELIMINARY; PRT; 390 AA.
AC Q8SR2;
DT 01-JUN-2002 (TEMBLrel. 21, Created)
DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
DE R565257P.
GN Names=CG5649;
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neuroptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nuncio J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY084184; AAL89922.1; -; mRNA.
DR FlyBase; FBgn0024887; CG5649.
DR FlyBase; FBgn0024887; kin17.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR007087; znf_C2H2.
DR SMART; SM00355; znf_C2H2_1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
KW Nuclear protein.
SQ
SEQUENCE 390 AA; 45286 MW; 39590CD3B61077CC CRC64;

Query Match 51.8%; Score 1050.5; DB 2; Length 390;
Best Local Similarity 52.2%; Pred. No. 1.2e-54;
Matches 206; Conservative 68; Mismatches 108; Indels 13; Gaps 5;

Qy 1 MGKSDFLSPKAIANRIKSGKGLQKLWYCOMCKQCRDENGPKCHCMSESHORQLLASEN 60
Db 1 MGRAEVGTPKYLANKMKSGKGLQKLWYCOMCKQCRDENGPKCHCMSESHORQLLPADN 60

Qy 61 PQQMDYFSEBEFRNDFLELLRRRTGTRVHNHNYVYNEISHREHNMATOWETLTPTK 120
Db 61 PGKFLHSPKSFSDGYMELLRRRTGTRVHNHNYVYNEISHREHNMATOWETLTPTK 120

Qy 121 WLGRGLCKVDTPKGVYQIVDRPETIRQLQLEKKKODLDEEKTAFIEQVRR-- 178
Db 121 WLGRGLCKVDTPKGVYQIVDRPETIRQLQLEKKKODLDEEKTAFIEQVRR-- 178

Qy 121 WLGRGLCKVDTPKGVYQIVDRPETIRQLQLEKKKODLDEEKTAFIEQVRR-- 180
Db 121 WLGRGLCKVDTPKGVYQIVDRPETIRQLQLEKKKODLDEEKTAFIEQVRR-- 180

Qy 179 RGLGKQEQE-TPVTFELSR-EENEKVTFFNFKGAGSAGATTSSKSLGPSALKLLGSA 236
Db 181 KAKGDEEGQKFTLKEENEPKLDIRLEK-----KFPDPTVLGKSALAKRPAP 232

Qy 237 ASGKRKSSQSAQAPAKKKKALDEIMELEBEKK-RTARTDAWLOPGIVVWIKYKLGEX 295
Db 233 EAEKVPKPKSVAGDSQTRSVLDEIIKQEBESKKEKRNKDYWLHKGIVVKFISKMGEX 292

Qy 296 YHKKGKGVKEVLDRTAVVKKMTSDGRLLKQDTHLETIVPAGKRVLVNGGYRGNEGL 355
Db 293 PFKQKAVVLDVDRYQGGIKFLETEKUKVQDAHLETIVPALDKPVMVNGAYRGSEALL 352

Qy 356 ESINEKAFSATVIETGPKLGRVVEGIQYEDISKL 390
Db 353 RKLDERYSVSVEILHGLPLKGRIVDNVQYEDISKL 387

```

```

RESULT 10
Q9ZVUS ARATH
ID Q9ZVUS ARATH PRELIMINARY; PRT; 411 AA.
AC Q9ZVUS;
DT 01-MAY-1999 (TEMBLrel. 10, Created)
DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
DT 01-FEB-2005 (TEMBLrel. 29, Last annotation update)
DE TSA14.13 protein (Hypothetical protein At1g55460).
GN Name=TSA14.13; Synonyms=At1g55460;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopses.
OX NCBI_TaxID=3702;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Pederspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA Altati H., Araujo R., Huizar L., Rowley D., Buehler E., Dunn P.,
RA Gonzalez A., Kremenetskaia I., Kim C., Lenz C., Li J., Liu S.,
RA Luros S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,
RA Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Etgu P., Lee J.M.,
RA Toriumi M., Yu G., Brooks S., Chao Q., Chen H., Karlin-Neumann G.,
RA Kim C., Lam B., Miranda M., Nguyen M., Palm C.J., Shinn P.,
RA Southwick A., Davis R.W., Ecker J.R., Theologis A.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005233; AAD10649.1; -; Genomic_DNA.
DR EMBL; AY051011; AAK93688.1; -; mRNA.
DR EMBL; AF360132; AAK25842.1; -; mRNA.
DR PIR; H96596; H96596.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR005824; KOW.
DR InterPro; IPR005824; KOW sub.
DR InterPro; IPR007087; znf_C2H2.
DR Pfam; PF00467; KOW; 1.
DR SMART; SM00739; KOW; 1.
DR SMART; SM00355; znf_C2H2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
KW Hypothetical protein.
SQ
SEQUENCE 411 AA; 47288 MW; 9DA6F8648002065D CRC64;

Query Match 48.7%; Score 988; DB 2; Length 411;
Best Local Similarity 48.8%; Pred. No. 7e-51;
Matches 203; Conservative 67; Mismatches 116; Indels 30; Gaps 7;

Qy 1 MGKSDFLSPKAIANRIKSGKGLQKLWYCOMCKQCRDENGPKCHCMSESHORQLLASEN 60
Db 1 MGRNDFITPKAIANRIKAKGLQKLWYCOMCKQCRDENGPKCHCMSESHORQVQCN 60

Qy 61 PQQMDYFSEBEFRNDFLELLRRRTGTRVHNHNYVYNEISHREHNMATOWETLTPTK 120
Db 61 PTRVVDGYSEFEQTFDLMRRSHRFSRIAATVYVYNEIYNDRHVHNMSTWATLTFFIK 120

Qy 121 WLGRGLCKVDTPKGVYQIVDRPETIRQLQLEKKKODLDEEKTAFIEQVRRG 180
Db 121 WLGRGLCKVDTPKGVYQIVDRPETIRQLQLEKKKODLDEEKTAFIEQVRRG 180

Qy 121 WLGRGLCKVDTPKGVYQIVDRPETIRQLQLEKKKODLDEEKTAFIEQVRRG 180
Db 121 WLGRGLCKVDTPKGVYQIVDRPETIRQLQLEKKKODLDEEKTAFIEQVRRG 180

```

```
QY 181 LE-----GKEQTPVFTSLRNEEB-----KYTFNLNKGAGGSAGATTSSK 222
Db 181 ASKLNKGSGEGGTSGNDEVDGDDDRKKXEDLRLKSGVKVGFAL---GGYKQVATGKB 237
QY 223 SSLGFSALKLLGSAAS-----GKRKSSQSAQPAKKKKSALDEIMLEEEKK-RTARTD 276
Db 238 R--GSSSKLLFGDEENDKVEREKRKRSKSGDSGRSEKERSALDELMKEEKKERMRKD 295
QY 277 AWLQPGIVVKIITKYLGEK-YHKHKGKGVKVEDRYTAVVMTDSGDRKLDQTHLETVP 335
Db 296 YWLFEGIIKVMKSKALAGYKQKGVKVKVDYVGEIKMLDSKHLVRVDQKELETVP 355
QY 336 ARKRVILNMGYRGNEGTLSEINAKFSATIVTETPLKGRVREGIQVEDISKLA 391
Db 356 QIGGMVKVINGAYRGSNARLLGVDTEKFCARQVQIEKGVVDGRVIRKSIYEDICKLA 411

RESULT 11
Q9XWF2 CAEEL PRELIMINARY; PRT; 404 AA.
AC Q9XWF2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-WAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein Y52B11A.9.
GN ORFNames=Y52B11A.9;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Bristol N2;
RX MEDLINE=9069613; PubMed=9851916; DOI=10.1126/science.282.5396.2012;
RG The C. elegans sequencing consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology.";
RL Science 282:2012-2018 (1998).
DR EMBL; AL032654; CAA21720.1; -; Genomic_DNA.
DR PIR; T27106; T27106.
DR WormBase; WBGene00013128; Y52B11A.9.
DR WormPep; Y52B11A.9; CE20297.
DR GO; GO:0003634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:007275; P:development; IEA.
DR InterPro; IPR002358; Ribosomal_L6_1.
DR InterPro; IPR007087; Znf_C2H2.
DR SMART; SM00355; Znf_C2H2; 1.
DR PROSITE; PS00525; RIBOSOMAL_L6_1; UNKNOWN 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 404 AA; 46401 MW; 9F55A1413525F057 CRC64;

Query Match 47.8%; Score 969.5; DB 2; Length 404;
Best Local Similarity 48.3%; Pred. No. 8.7e-50;
Matches 199; Conservative 75; Mismatches 109; Indels 29; Gaps 8;

QY 1 MGKSDFLSPKATNRITKSGLOKLRVYCOMCQKQCRDENGFKCHCWSESHOROLLILAS 60
Db 1 MGKHEGSKDLNARITKSGLOKLRVYCOMCQKQCRDANGFKCHLTSEAHORQLLFAEN 60
QY 61 PQQFMDYFSEEFNDLFLRRRFGKRVHNNIVYNEYISHREIHIMNATQWETLDFTK 120
Db 61 SNSYLQFSDNDFEKNFMQLRLTSYGYKRVANEVYNAFVKDGHVHVNSTVWHSLTGFVQ 120
QY 121 WLGREGLCKVDTPKGYWYQYIDRDPETTRRQLELEKKKKQDLDDSEKTAFTIEQVRG 180
Db 121 YLGSSGCKKIDEGKGYWYAYIDQ--EALIRKEEDQRQKQKDDDERHMQMDGMVQRG 178
QY 181 --LEGKEQTPVFTSLRNEEBEKVTFNLNKG-----AGGSAGATTSSKSSL----- 225
```

```
Db 179 KELAGDDHEYEATELIRDTDPQKIQLDLNLGILDRKLDVKSGVASAKISIFDMPKVKE 238
QY 236 -----GPSALKLGSAAAGKRSQSSQAQPAKKKKSALDIMELEEBKK-RTARTDAWL 279
Db 239 DPDEPGPSQ-----PSRKSGKRSRSRSPAANKFKKSALDEIKEMEERKKERKNRKYWM 294
QY 280 QPGIWKIITKYLGEKHYHKGKGVKVEDRYTAVVMTDSGDRKLDQTHLETVIPAPGK 339
Db 295 REGIVKVVITSLGSEYYKAGVVRKVVDDYTAQVKL--DGGTVVKLDQEHVETVIPSLO 353
QY 340 RVLNMGYRGNEGTLSEINAKFSATIVTETPLKGRVREGIQVEDISKLA 391
Db 354 QMMIVNGAYRQEQATLESIDEKFTSLRLKIASGPTGRQID-VPYEDASKLA 404

RESULT 12
Q4TDV9 TETNG PRELIMINARY; PRT; 269 AA.
AC Q4TDV9;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Chromosome undetermined SCAF6005, whole genome shotgun sequence.
DE (Fragment).
GN ORFNames=GSTENG0002601001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
Mauceli E., Bounneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
Dasilva C., Salanoubat M., Levy M., Boudet N., Castellan S.,
Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
Parra G., Lardier G., Chapelle C., McKernan K.J., McEwan P., Bosak S.,
Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
Winkler P., Lander E.S., Weissbach J., Roest Crollius H.;
RA Craud C., Duprat S., Brottier C., Coutanceau J.P., Gouzy J.,
RA Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
Winkler P., Lander E.S., Weissbach J., Roest Crollius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
the early vertebrate proto-karyotype.";
RL Nature 431:946-957 (2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
preliminary data.
CC EMBL; CAE01006005; CAF88923.1; -; Genomic_DNA.
DR NON_TER 1
FT NON_TER 269
SQ SEQUENCE 269 AA; 31736 MW; E6883839A2CD17AF CRC64;

Query Match 46.0%; Score 932.5; DB 2; Length 269;
Best Local Similarity 68.4%; Pred. No. 8.6e-48;
Matches 186; Conservative 33; Mismatches 38; Indels 15; Gaps 5;

QY 55 LLAENPQQFMDYFSEEFNDLFLRRRFGKRVHNNIVYNEYISHREIHIMNATQWET 114
Db 1 LLAENPNRFDHDFQEFKQFLELRRRFGKRVHNNIYNEYISDRQIHIMNATRWET 60
QY 115 LTDFTKWLGREGLCKVDTPKGYWYQYIDRDPETTRRQLELEKKKKQDLDDSEKTAFTIE 174
Db 61 LTDFTKWLGREGLCKVDTPKGYWYQYIDRDPETTRRQLELEKKKKQDLDDSEKTAFTIE 120
QY 175 EQVRGLEGKE-QETPVFTSLRNEEBEKVTFNLNKGAGGSAGATTSSKSLGPSALKLL 233
```

	: : : : : :	
Db	121 EQVRRGRDRETEETPLYTELKRNEBEKFAFN--GASLCVAGPSKSPAVGVGALKE	178
Qy	234 GSAAGSKRKSSQAQPAKKKKAIDIMELEBK---TARTDAWLQPGLVVKVIITK	290
Db	179 AAAASVRKDAS-SESSETKRKRKSAIIEIMEEKKKQQQLRTDYWLQPNLVVKVIITK	237
Qy	291 KLGEKYHKKGKVVEVDTRYTVAVMTDSGR	322
Db	238 KLGERYHKKRGVV-----MVRPRDSVDR	261

RESULT 13

Q75LU5_ORYSA	PRELIMINARY;	PRT;	430 AA.
ID	Q75LU5_ORYSA		
AC	Q75LU5;		
DT	05-JUL-2004	(TrEMBLrel. 27, Created)	
DT	05-JUL-2004	(TrEMBLrel. 27, Last sequence update)	
DT	05-JUL-2004	(TrEMBLrel. 27, Last annotation update)	
DE	Hypothetical protein OSJNB0015121.3.		
DE	Name=OSJNB0015121.3;		
GN	Oryza sativa (japonica cultivar-group).		
OS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophytes; Magnoliophyta; Liliopsida; Poales; Poaceae;		
OC	Ehrhartoideae; Oryzoae; Oryza.		
NCBI_TaxID=39947;			
[1]			
RN	NUCLEOTIDE SEQUENCE.		
RP	Buell C.R., Yuan Q., Quyang S., Liu J., Gansberger K., Jones K.M.,		
RA	Overton II L.L., Tseltrin T., Kim M.M., Bera J.J., Jin S.S.,		
RA	Fadrosch D.W., Tallon L.J., Koo H., Zismann V., Hsiao J., Blunt S.,		
RA	Vanaken S., Riedmuller S.B., Uterback T.T., Feldblyum T.V.,		
RA	Yang Q.Q., Haas B.J., Suh B.B., Peterson J.J., Quackenbush J.,		
RA	White O., Salzberg S.L., Fraser C.M.;		
RA	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.		
RL	[2]		
RN	NUCLEOTIDE SEQUENCE.		
RP	Buell R.;		
RA	Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.		
RL	EMBL; AC091302; AAR00634.1; -; Genomic_DNA.		
DR	Gramene; Q75LU5; -.		
DR	InterPro; IPR005824; KOW.		
DR	InterPro; IPR007087; Znf_C2H2.		
DR	Pfam; PF00467; KOW; 1.		
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.		
DR	Hypothetical protein.		
KW	SEQUENCE 430 AA; 49067 MW; 6151D748BD2F4690 CRC64;		
SO			

Db	299	MKEEKAKERSNRKQWLCFGLVVKVMSLAEKGYCKQGVVKKRVIDKYVGEIEMLESK	358
QY	321	DRLLQDQTHLETVIPAPGKRVLVLVANGVYRGNGSTLSEINEKAFSATIVITETGPLKGRRVE	380
Db	359	HVLRVDQDELETVIPQIGLVRIVANGAVRGSNARLLSLVDTERFCAKQVQVEKGLYDGKVLK	418
QY	381	GIQYEDISKL 390	
Db	419	AIQYEDICK1 428	
RESULT 14			
Q9AY87_ORYSA			
ID	Q9AY87_ORYSA	PRELIMINARY;	PRT; 402 AA.
AC	Q9AY87;		
DT	01-JUN-2001	(TEMBLrel. 17, Created)	
DT	01-JUN-2001	(TEMBLrel. 17, Last sequence update)	
DT	01-MAR-2004	(TEMBLrel. 26, Last annotation update)	
DE	Hypotheical protein OSJNBA0004B24.17.		
GN	Name=OSJNBA004B24.17;		
OS	Oryza sativa (Rice).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;		
OC	Erihartoideae; Oryzeae; Oryza.		
OX	NCBI_TaxID=4530;		
RN	[1]		
RP	NUCLEOTIDE SEQUENCE.		
RA	Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Burr P.C., Hsiao J.,		
RA	Ziemann V., Pai G., Bowman C.L., Fujii C.Y., VanAken S.E.,		
RA	Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblyum T.V.,		
RA	Quackenbush J., White O., Salzberg S.L., Fraser C.M.;		
RL	Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	NUCLEOTIDE SEQUENCE.		
RA	Buell R.;		
RL	Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL:	AC084319; AAG59654.1; -; Genomic_DNA.	
DR	Gramene;	Q9AY87; -	
DR	GO;	GO:0005634; C:nucleus; IEA.	
DR	GO;	GO:0046872; F:metal ion binding; IEA.	
DR	GO;	GO:0003676; F:nucleic acid binding; IEA.	
DR	GO;	GO:0008270; F:zinc ion binding; IEA.	
DR	InterPro;	IPR007087; Znf_C2H2.	
DR	SMART;	SM00355; Znf_C2H2_1	
DR	PROSITE;	PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.	
DR	Hypotheical protein.		
SQL	SEQUENCE 402 AA; 45962 MW; A1FE38BCC60D13E CRC64;		


```

; ORGANISM: Homo sapiens
US-10-144-198-4
Query Match      6.0%; Score 122; DB 2; Length 1564;
Best Local Similarity 24.2%; Pred. No. 0.077;
Matches 66; Conservative 38; Mismatches 91; Indels 78; Gaps 13;

QY 116 TPTFKWLGREGKCKVDET-----PKG-WYQYIDRDPETIRROL-ELEKKKKQDLDD 165
DB 40 TQCRNWL-KTGNCLYNTCRFVHGSPRGKGYSSNRYRSPERTGDLRMRMKNKRDQDVT 98

QY 166 E-----EKTAKFIEQVRGLEGKEQETPVFTLSRENEEEKVTFLNKGAGGSAG--- 216
DB 99 EPQKRNTESSSPVKRESSGRHREKEDIKITERTPESEEEENVEVETNRDSDNGDINY 158

QY 217 -----ATTSKSSSLGPSALKLLGSA 237
DB 159 DYVHELSELMKROKIQRELMKLEQENNEKREIIKKVEPSVVRVSKLSPSP-SLRKSK 217

QY 238 SGKRESSOSSAQPAKKKSA-----LDE--IMELEEEKKRTARTDAMLQRCIVVKIIT 289
DB 218 SPKRKSPKSSASKDKRTSAVSPLLDQQRNSKTNOSKKGPRTPS--PPPPPIEDIA 275

QY 290 KILGEKYHKKGVVVVIDRYTAVVVKMTDSGDR 322
DB 276 --LGKKY-KKKYKVKDRIEB-----KTRDGKDR 300

RESULT 6
US-09-824-574-7
; Sequence 7, Application US/09824574
; Patent No. 6693170
; GENERAL INFORMATION:
; APPLICANT: Rouleau, Natalie
; APPLICANT: Moilanen, Anu-Maarit
; APPLICANT: Palvimo, Jorma J.
; APPLICANT: Jone, Olli A.
; TITLE OF INVENTION: ARI4 Gene and Protein
; FILE REFERENCE: 2630-109
; CURRENT APPLICATION NUMBER: US/09/824,574
; CURRENT FILING DATE: 2001-04-03
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 7
; LENGTH: 2476
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-824-574-7

Query Match      6.0%; Score 121.5; DB 2; Length 2476;
Best Local Similarity 22.8%; Pred. No. 0.17;
Matches 68; Conservative 41; Mismatches 108; Indels 81; Gaps 13;

QY 122 LGREGKCKVDETPKGWYIQYIDRDP-----ETIRROLE--LEKKKKQDLDDDEKTKAK 171
DB 247 LGRKELSTIMDENQWYCYICQPEFLDLVTACNSVFNLEQLQONKKIKVDSEKTSK 306

QY 172 FIEQVRGLEGKEQETPVFTLSRENEEEKVTFLNKGAGGSAGATTYSK----- 222
DB 307 VCDQ-----TSKFSPPKSSSCNGEKK-----LEESCGSVSYSHSALSVPKEMI 354

QY 223 -----SSLGPSALKLIGSAASKRESSOSSAQPAKKKYSALDEI-----MELEEE 268
DB 355 KKTTLKLIETTSNNSSVIKFLQAADNSEMTSAMKLCQ-LKSPKSVLDDIKKAHLEED 413

QY 269 KRTARTDAMLQRCIVVKIITKILGEKYHKKGVVVVIDRYTAVVVKMTDSGDRKLQDT 328
DB 414 LNSEIQ-----LDDVHKEKN-----TKDLKSTDAKSTKLKGK 447

QY 329 HLETVIPAPGRKRVLVNGVGRNEGTLNESINEKAFSATIVETGLPKGR-RVEGIQYE 385
DB 448 --EKSYSYTERKEEFLKLD--RSSVKAIDGEEQRAHKST-----SCEHGSGRKGDSQYE 497
```

```

RESULT 7
US-08-285-440-6
; Sequence 6, Application US/08285440
; Patent No. 5532337
; GENERAL INFORMATION:
; APPLICANT: Ken'ichiro HAYASHI et al.
; TITLE OF INVENTION: POLYPEPTIDE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: DisplayWrite
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/285,440
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/858,947
; FILING DATE: March 27, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 558 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEetical:
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM:
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE:
; CELL LINE:
; ORGANELLE:
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE:
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; MAP POSITION:
; UNITS:
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; PUBLICATION INFORMATION:
; AUTHORS:
; TITLE:
; JOURNAL:
; VOLUME:
```


ISSUE: 558 amino acids
PAGES: amino acid
DATE: STRANDEDNESS: single
DOCUMENT NUMBER: TOPOLOGY: linear
FILING DATE: MOLECULE TYPE: peptide
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-08-285-440-6

Query Match 6.0%; Score 121; DB 1; Length 558;
Best Local Similarity 22.8%; Pred. No. 0.02;
Matches 74; Conservative 45; Mismatches 123; Indels 82; Gaps 14;

QY 48 ESHQQLLASNPQOFMDYFSEEFNDPFLLELRRTGKRVHNNIVVNEYISHREHIHM 107
DB 99 ERRQRKQALERQKE-----FDPITTDASLSLPSRRMQNDTAENETTEKEEKSESQRQERY 154
QY 108 NATOWETLT-DFTK--WLGREGLCQVD-----ETPKGWYIQYIDRDPETIRQLELE 156
DB 155 EIEETETVTKSYQKNDWRDAENKKEDKEKEEBEKKRGSIGENQGEKGTQVQAKRE 214
QY 157 KK-----KKQDLDEE-KTAKFIEEQVRRGLEKGEQETPVFTLSRENEE-----EKV 203
DB 215 KLQEDKPTFKKEIKDEKIKKDKPEKVKSFMDRKG---FTEVKSQNGEFMTHKLKH 270
QY 204 TPNLKGAGGSAGATTSSKSSSLGPSALKLGSAAAGKR-----KES 244
DB 271 TENTFSRPGGRASVDTKEAG-----APQVEAGKRLLELRRTGGETESEFEKLKQK 322
QY 245 SQSSA---QPAKKKSALDEIMLEEEKKRTARTDQWLPQGVVVKIITKLGKGYHKKKG 301
DB 323 QQEALELEELKKREERRKVLLEEEQRRKQBEAD-----RKUREEKKRR- 368

QY 302 VVKEVIDRYTAVV-----KMTDSG 320
DB 369 -LKEEIERRAEAAEKQKMPEDG 391

RESULT 8
US-08-630-349-6
Sequence 6, Application US/08630349
Patent No. 5739008
GENERAL INFORMATION:
APPLICANT: Ken'ichiro HAYASHI et al.
TITLE OF INVENTION: POLYPEPTIDE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESS: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 144 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,349
FILING DATE: April 10, 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/285,440
FILING DATE: August 4, 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/858,947
FILING DATE: March 27, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:

LENGTH: 558 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-630-349-6

Query Match 6.0%; Score 121; DB 1; Length 558;
Best Local Similarity 22.8%; Pred. No. 0.02;
Matches 74; Conservative 45; Mismatches 123; Indels 82; Gaps 14;

QY 48 ESHQQLLASNPQOFMDYFSEEFNDPFLLELRRTGKRVHNNIVVNEYISHREHIHM 107
DB 99 ERRQRKQALERQKE-----FDPITTDASLSLPSRRMQNDTAENETTEKEEKSESQRQERY 154
QY 108 NATOWETLT-DFTK--WLGREGLCQVD-----ETPKGWYIQYIDRDPETIRQLELE 156
DB 155 EIEETETVTKSYQKNDWRDAENKKEDKEKEEBEKKRGSIGENQGEKGTQVQAKRE 214
QY 157 KK-----KKQDLDEE-KTAKFIEEQVRRGLEKGEQETPVFTLSRENEE-----EKV 203
DB 215 KLQEDKPTFKKEIKDEKIKKDKPEKVKSFMDRKG---FTEVKSQNGEFMTHKLKH 270
QY 204 TPNLKGAGGSAGATTSSKSSSLGPSALKLGSAAAGKR-----KES 244
DB 271 TENTFSRPGGRASVDTKEAG-----APQVEAGKRLLELRRTGGETESEFEKLKQK 322
QY 245 SQSSA---QPAKKKSALDEIMLEEEKKRTARTDQWLPQGVVVKIITKLGKGYHKKKG 301
DB 323 QQEALELEELKKREERRKVLLEEEQRRKQBEAD-----RKUREEKKRR- 368

QY 302 VVKEVIDRYTAVV-----KMTDSG 320
DB 369 -LKEEIERRAEAAEKQKMPEDG 391

RESULT 9
US-09-949-016-7610
Sequence 7610, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7610
LENGTH: 1676
TYPE: PRT
ORGANISM: Human
US-09-949-016-7610

Query Match 5.9%; Score 119.5; DB 2; Length 1676;
Best Local Similarity 20.2%; Pred. No. 0.15;
Matches 84; Conservative 50; Mismatches 113; Indels 169; Gaps 16;

QY 2 GKSDFLSPKATANRIKSGLOKLRWYCOMCKQCRDENGFKCHCMSESHQRLLASNP 61
DB 506 GKHQIQISRLQDVRLLK-KQTQKTE--LEVLKQCDLE-----IWEIKLQOQL----- 550
QY 62 QQFMDYFSEEFNDPFL-----ELRRRF-----GTRVHNNIVVNEYISHREH 104
DB 551 -----QEYQNKLIYLVPEKQLINERIKNMQFSNTPDGVSLLHKKLSKEELCQRLK 602


```
QY 266 EEEKGTARTDAWLQPGIIVKIIITKLGKGYHKHKKGVVKE 305
Db 1254 EEEKR-----IEAEKKABEERKRIBAVKK 1278

RESULT 12
US-09-949-016-8999
; Sequence 8999, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8999
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8999

Query Match 5.8%; Score 117; DB 2; Length 484;
Best Local Similarity 20.2%; Pred. No. 0.039;
Matches 67; Conservative 48; Mismatches 93; Indels 124; Gaps 14;

QY 48 ESHQROLLENSENPOQFMDYFSEEFNDLELLRRRFGTKRVHNNIVYNEYISHREHIH 107
Db 108 ERRQKRLQALEREKGFDTITDASL-----LPSRRMQNDTAENETTEKEEKSES 158
QY 108 NATQWETLDTFTKWLREGGLCKVDETPKGWYIQYIDRDPETIRRQ-----ELEKKKQ 161
Db 159 RQERYE-----IEET-----ETVTKSQKNDWRDAENKKE 189
QY 162 DLDEKTKAKTIEEQVRRLG-----KEQETPV-----FTLSRENEE 200
Db 190 DKEKESEE---EEKPKRGSIGENQIKDEKIKKDEPKBEVKSFMDRKKGFTVKSQNGE 245
QY 201 -----EKVTNFKGAGGAGATTSSSLGPSALKLLGSAASGKR----- 241
Db 246 FMTHKLKHTENTFSRPGGRASVDTKAEG-----APQVEAGKRLBELRRRRRGETESE 297
QY 242 -----KESQSSA---QPAKKKSALDEMELEEKKRTARTDAWLQPGIIVKIIITKLG 293
Db 298 EPEKLKQKQEALEELKKEERKRVLEEEQRRKQBEAD-----RKL 344
QY 294 EKYHKKGKGVVKEVIDRYTAVV-----KMTDSG 320
Db 345 EEEKRR--LKEBIERRRAEAEKQKMPEDG 374

RESULT 13
US-08-285-440-5
; Sequence 5, Application US/08285440
; Patent No. 5532337
; GENERAL INFORMATION:
; APPLICANT: Ken'ichiro HAYASHI et al.
; TITLE OF INVENTION: POLYPEPTIDE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
```

```
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: DisplayWrite
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/285,440
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/858,947
; FILING DATE: March 27, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 532 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL:
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM:
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL LINE:
; ORGANELLE:
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE:
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; MAP POSITION:
; UNITS:
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; PUBLICATION INFORMATION:
; AUTHORS:
; TITLE:
; JOURNAL:
; VOLUME:
; ISSUE:
; PAGES:
; DATE:
; DOCUMENT NUMBER:
; FILING DATE:
; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO:
US-08-285-440-5

Query Match 5.8%; Score 117; DB 1; Length 532;
Best Local Similarity 20.2%; Pred. No. 0.045;
Matches 67; Conservative 48; Mismatches 93; Indels 124; Gaps 14;
```

QY

```

Db 99 ERQKRLQEALEKQKFDPTITDASLS-----LPSRRQNDTAENETTEKEEKSES 149
Qy 108 NATOMETLTDFTKWLREGGLCKVDETPKGWIQYIDRDPETIRROL-----ELEKXKQ 161
Db 150 RQERYE-----ETVTKSYQKNDWRDAEENKKE 180
Qy 162 DLDEEKTAKFIEQVVRGLEG-----KEQETPV-----FTLSRENEE 200
Db 181 DKEKEEEE-----BEKPKRGSGIGENQIKDEKIKDKEPKKEVKFMDRKKGFTVKSQNGE 236
Qy 201 -----EKVTFNLNKGAGGSAGATTSSKSSSLGPSALKLLGSAASGKR----- 241
Db 237 FMTHKLKHTENTSRPGGRASVDTKEAG-----APQVEAGKRLLEELRRRGETESE 288
Qy 242 -----KESSOSSA---OPAKKKKSALDEIMELBEEKKRTARTDAWLOPGIVVVIITKKLG 293
Db 289 EFEKLKQKQQAALAELEELKKEERKRVLEEEBQRRKQBEAD-----RKLR 335
Qy 294 EKYHKKGVVKEVIDRYTAVV-----KMTDSG 320
Db 336 EBEKRR--LKBEIERRRRAEAERKQKMPEDG 365

```

RESULT 14

```

US-08-630-349-5
; Sequence 5, Application US/08630349
; Patent No. 5739008
; GENERAL INFORMATION:
; APPLICANT: Ken'ichiro HAYASHI et al.
; TITLE OF INVENTION: POLYPEPTIDE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 144 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,349
; FILING DATE: April 10, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/285,440
; FILING DATE: August 4, 1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/858,947
; FILING DATE: March 27, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 532 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-630-349-5

```

```

Query Match 5.8%; Score 117; DB 1; Length 532;
Best Local Similarity 20.2%; Pred. No. 0.045;
Matches 67; Conservative 48; Mismatches 93; Indels 124; Gaps 14;

Qy 48 ESHQRQLLASENPOQFMDYFSEFRNDFLELLRRRFGTKRVHNNIVVYNEISHREIHM 107
Db 99 ERQKRLQEALEKQKFDPTITDASLS-----LPSRRQNDTAENETTEKEEKSES 149

```

```

Qy 108 NATOMETLTDFTKWLREGGLCKVDETPKGWIQYIDRDPETIRROL-----ELEKXKQ 161
Db 150 RQERYE-----ETVTKSYQKNDWRDAEENKKE 180
Qy 162 DLDEEKTAKFIEQVVRGLEG-----KEQETPV-----FTLSRENEE 200
Db 181 DKEKEEEE-----BEKPKRGSGIGENQIKDEKIKDKEPKKEVKFMDRKKGFTVKSQNGE 236
Qy 201 -----EKVTFNLNKGAGGSAGATTSSKSSSLGPSALKLLGSAASGKR----- 241
Db 237 FMTHKLKHTENTSRPGGRASVDTKEAG-----APQVEAGKRLLEELRRRGETESE 288
Qy 242 -----KESSOSSA---OPAKKKKSALDEIMELBEEKKRTARTDAWLOPGIVVVIITKKLG 293
Db 289 EFEKLKQKQQAALAELEELKKEERKRVLEEEBQRRKQBEAD-----RKLR 335
Qy 294 EKYHKKGVVKEVIDRYTAVV-----KMTDSG 320
Db 336 EBEKRR--LKBEIERRRRAEAERKQKMPEDG 365

```

RESULT 15

```

US-09-949-016-6479
; Sequence 6479, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6479
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6479

```

```

Query Match 5.8%; Score 117; DB 2; Length 538;
Best Local Similarity 20.2%; Pred. No. 0.045;
Matches 67; Conservative 48; Mismatches 93; Indels 124; Gaps 14;

Qy 48 ESHQRQLLASENPOQFMDYFSEFRNDFLELLRRRFGTKRVHNNIVVYNEISHREIHM 107
Db 105 ERQKRLQEALEKQKFDPTITDASLS-----LPSRRQNDTAENETTEKEEKSES 155
Qy 108 NATOMETLTDFTKWLREGGLCKVDETPKGWIQYIDRDPETIRROL-----ELEKXKQ 161
Db 156 RQERYE-----ETVTKSYQKNDWRDAEENKKE 186
Qy 162 DLDEEKTAKFIEQVVRGLEG-----KEQETPV-----FTLSRENEE 200
Db 187 DKEKEEEE-----BEKPKRGSGIGENQIKDEKIKDKEPKKEVKFMDRKKGFTVKSQNGE 242
Qy 201 -----EKVTFNLNKGAGGSAGATTSSKSSSLGPSALKLLGSAASGKR----- 241
Db 243 FMTHKLKHTENTSRPGGRASVDTKEAG-----APQVEAGKRLLEELRRRGETESE 294
Qy 242 -----KESSOSSA---OPAKKKKSALDEIMELBEEKKRTARTDAWLOPGIVVVIITKKLG 293
Db 295 EFEKLKQKQQAALAELEELKKEERKRVLEEEBQRRKQBEAD-----RKLR 341
Qy 294 EKYHKKGVVKEVIDRYTAVV-----KMTDSG 320

```

Db 342 EEEKER--LKEIERRRAAEKROKMPDG 371

Search completed: November 23, 2005, 16:35:41
Job time : 35.9133 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:24:04 ; Search time 113.709 Seconds
(without alignments)
1436.746 Million cell updates/sec

Title: US-09-555-529-25
Perfect score: 2028
Sequence: 1 MGKSDFLSPKAIANRIKSG.....GPKLGRVBEIQVEDISKLA 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA Main:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pap.*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pap.*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pap.*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pap.*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pap.*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1895	93.4	393	4	US-10-408-765A-1465
2	1052.5	51.9	390	6	US-11-097-143-41040
3	913	45.0	437	4	US-10-425-114-48183
4	913	45.0	437	4	US-10-425-114-58982
5	908	44.8	424	4	US-10-425-115-341788
6	901.5	44.5	423	5	US-10-739-930-8499
7	613	30.2	176	4	US-10-424-599-168326
8	531	26.2	138	4	US-10-437-963-191916
9	449	22.1	120	4	US-10-767-701-57151
10	361	17.8	84	4	US-10-106-698-5464
11	324	16.0	162	4	US-10-424-599-196166
12	239	11.8	123	4	US-10-425-115-300949
13	221.5	10.9	107	4	US-10-425-115-230934
14	209	10.3	109	4	US-10-424-599-196168
15	149	7.3	278	4	US-10-425-114-46572
16	149	7.3	430	4	US-10-425-115-284883
17	142	7.0	381	4	US-10-437-963-194407
18	142	7.0	428	4	US-10-437-963-194408
19	134	6.6	375	5	US-10-739-930-6249
20	133.5	6.6	2052	4	US-10-437-963-137285
21	130.5	6.4	1212	5	US-10-788-793-2
22	127	6.3	1359	5	US-10-732-923-8707
23	127	6.3	1359	5	US-10-732-923-8708
24	126.5	6.2	785	4	US-10-425-115-332751
25	125.5	6.2	158	4	US-10-425-115-241450
26	124.5	6.1	1080	4	US-10-369-493-12560
27	124	6.1	726	5	US-10-483-680-2

ALIGNMENTS

RESULT 1

US-10-408-765A-1465
; Sequence 1465, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:

; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Bojin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.

; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; IDENTIFIED IN THE MITOCHONDRIAL PROTEOME

; FILE REFERENCE: 660088.465

; CURRENT APPLICATION NUMBER: US/10/408,765A

; CURRENT FILING DATE: 2003-04-04

; NUMBER OF SEQ ID NOS: 3077

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1465

; LENGTH: 393

; TYPE: PRT

; ORGANISM: Homo sapiens

; US-10-408-765A-1465

Query Match 93.4%; Score 1895; DB 4; Length 393;
Best Local Similarity 92.4%; Pred. No. 1.1e-137;
Matches 363; Conservative 14; Mismatches 14; Indels 2; Gaps 1;

Qy	1	MGKSDFLSPKAIANRIKSGKQLKRWYQCMQCRDENGFKCHMSHQRQLLASEN	60
Db	1	MGKSDFLTPKAIANRIKSGKQLKRWYQCMQCRDENGFKCHMSHQRQLLASEN	60
Qy	61	PQOFMDYFSEFPNDLFELRRRFGTKRVNNIYVYNIYSHRHHNNATOWETLTDTPTK	120
Db	61	PQOFMDYFSEFPNDLFELRRRFGTKRVNNIYVYNIYSHRHHNNATOWETLTDTPTK	120
Qy	121	WLGRGLCKVDPTPKGWYQIYIDRDPTIRRLQLEKKKQDLDEEKTAKFIEQVRRG	180
Db	121	WLGRGLCKVDPTPKGWYQIYIDRDPTIRRLQLEKKKQDLDEEKTAKFIEQVRRG	180
Qy	181	LEGKEQTPVFTLSRNEBEKVTFNUNKAGGAGATTSSSLGFSALKLSAASGK	240
Db	181	LEGKEQVPTFTLSRENDEBEKVTFNLSKGACSSSGATSSKSTLGFSAKLTIGSSASVK	240
Qy	241	RKSSQSSAQ--PAKKKKSALDEIMELEEEKKRTARTDAMLQPGIVVKITIKLGEKYHK	298
Db	241	RKSSQSSQSTOSKEKKKKKKSALDEIMELEEEKKRTARTDYLQPEIIVKIIITKLGEKYHK	300
Qy	299	KGVVKEVIDRYTAVVKNMTSDGRLKLDQTHLETVIPAPGKRVLVLNGVYRNGEGLTLESI	358

Sequence 50637, A
Sequence 50638, A
Sequence 39688, A
Sequence 2, Appl1
Sequence 4, Appl1
Sequence 6, Appl1
Sequence 2119, Ap
Sequence 7, Appl1
Sequence 8136, Ap
Sequence 35, Appl1
Sequence 332760, A
Sequence 56908, A
Sequence 152, App
Sequence 48614, A
Sequence 176714, A
Sequence 16, Appl
Sequence 3309, Ap
Sequence 41636, A

28 124 6.1 2400 5 US-10-450-763-50637
29 124 6.1 2415 5 US-10-450-763-50638
30 122 6.0 435 5 US-10-450-763-39688
31 122 6.0 1564 4 US-10-144-198-2
32 122 6.0 1564 4 US-10-144-198-4
33 121.5 6.0 1114 4 US-10-408-765A-2119
34 121.5 6.0 1213 5 US-10-788-793-6
35 121.5 6.0 2476 3 US-09-824-574-7
36 121.5 6.0 2476 5 US-10-732-923-8136
37 121 6.0 439 5 US-10-805-684-35
38 120.5 5.9 911 4 US-10-425-115-332760
39 120.5 5.9 1156 5 US-10-450-763-56908
40 120 5.9 805 4 US-10-369-493-152
41 120 5.9 856 4 US-10-282-122A-48614
42 120 5.9 1382 4 US-10-437-963-176714
43 119.5 5.9 1681 4 US-10-398-885A-16
44 119 5.9 978 5 US-10-732-923-3309
45 119 5.9 1080 5 US-10-450-763-41636

Db 301 KKAIVKEVDKTAVVVMDSGDKLDDQTHLETVIPAKGRILVNLGNGYRGNEGTLSEI 360
QY 359 NEKAFSATIVETGPKLGRVREGIOYEDISKLA 391
Db 361 NEKTFSATIVETGPKLGRVREGIOYEDISKLA 393

RESULT 2
US-11-097-143-41040
; Sequence 41040, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41040
; LENGTH: 390
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-41040

Query Match 51.9%; Score 1052.5; DB 6; Length 390;
Best Local Similarity 52.2%; Pred. No. 1.1e-72;
Matches 206; Conservative 69; Mismatches 107; Indels 13; Gaps 5;
QY 1 MGKSDFLSPKAIANRIKSKGLQKLRYWCQKQCRDENGFKCHCMSESHQROLILASEN 60
Db 1 MGRAEVTGPKYLANKMKSKGLQKLRYWCQKQCRDENGFKCHCMSESHQROLILFADN 60
QY 61 PQOFMDYFSEEFNRDLELLRRFGRVHNNIVYNEIYSHREHIMHNATQETLTDFTK 120
Db 61 PQKFLHSFKSEFSDGFWELLRRRFGFKRTSANKIYOETIAHKEHIMHNATRLTSLSDYVK 120
QY 121 WLGRGLCKVDTPKGYIYQIDRDPETIRROLELEKKKKQDLDDEKTAKEIEQVR-- 178
Db 121 WLGRGTQVIADETEKGMFWYIDRSPAEAMERQAKADKEKMEKDDERWADFIEQIQKNA 180
QY 179 RLGEKQE--TPVFTLSR-ENEEKVTNLNKGAGSGAGATTSKSSSGPSALKLLGSA 236
Db 181 KAKDGEDEGQBKFTBLKRENEPLKDIRLEK-----KFPDPTVLGKSALAKRPAP 232
QY 237 ASGRKESQSSQAQPAKKKKSALDEMELEKKK-RTARTDAWLQGIYVVKIITKKLGEK 295
Db 233 EAEKVFKKPKSVAGDSQTRSVDELIIKQEEKKERANRKYVWLHGIYVVKFTSKSMGEK 292
QY 296 YHKKGKVVKEVDRTYAVVVMKTDGDRLLKLDQTHLETVIPAGKRVLVNLGNGYRGNEGTL 355
Db 293 FFKQKAVLVDIDRYOGKIKFLETGKVKLDQAHLETVIPALDKPVMVNGAYRGSEALL 352

QY 356 ESINEKAFSATIVETGPKLGRVREGIOYEDISKL 390
Db 353 RKLDBERRYSVSEILHLGPKLGRIVDNVQYEDISKL 387
RESULT 3
US-10-425-114-48183
; Sequence 48183, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 48183
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3591-011-E3_FLI.pep
US-10-425-114-48183

Query Match 45.0%; Score 913; DB 4; Length 437;
Best Local Similarity 44.9%; Pred. No. 7.3e-62;
Matches 190; Conservative 75; Mismatches 126; Indels 32; Gaps 7;
QY 1 MGKSDFLSPKAIANRIKSKGLQKLRYWCQKQCRDENGFKCHCMSESHQROLILASEN 60
Db 14 MGKEFLTPKAIANRIKAKGLQKLRYWCQKQCRDENGFKCHCMSESHQRMQVFGWA 73
QY 61 PQOFMDYFSEEFNRDLELLRRRFGFKRVHNNIVYNEIYSHREHIMHNATQETLTDFTK 120
Db 74 PDRVVEGFSEFLESFLSLIRRAHRSVAATVVYNEIADRRHHVHMNSTRWATLTFYVK 133
QY 121 WLGRGLCKVDTPKGYIYQIDRDPETIRROLELEKKKKQDLDDEKTAKEIEQVR- 179
Db 134 FLGREGYCKVEDTPKGMFTYIDRSEQAVERLKRKRKIKSDMADDERQERMIARQIERA 193
QY 180 -----GLEGEQETPVFTLSRENEBEKVTFNLNKGAGGSAG--ATTSKSSSLQP 227
Db 194 HKSLAKPNNGGNAEGEPESGSEGEYSGSDDEQEBEDDSKEADKATGKIAIALQKAVGP 253
QY 228 SA-----LKL-----LGSAAAGKRK-----ESSQSSAQAQPAKKKKSALDEMELEEE- 269
Db 254 KVNPPDDPKPKKFGPDEEDGSGTRDQEKNELTKMGKDVKAEEAKRSALDELMKEEEKAK 313
QY 270 KRTARTDAWLQGIYVVKIITKKLGEK-YHKKGKVVKEVIDRYTAVVVMKTDGDRLLKDOT 328
Db 314 ERSNFKDQWLCGIVVVKMSKSLAEKGYKQKGVVKKVVDKYVGEIEMDESXHVLRVDDQ 373
QY 329 HLETVIPAGKRVLVNLGNGYRGNEGTLSEINEKAFSATIVETGPKLGRVREGIOYEDIS 388
Db 374 ELETVIPQIGGLVIRVNGAYRGSNARLLSVDTKEFCAKVQVEKGLYDGVKVLRAVEYEDIC 433
QY 389 KLA 391
Db 434 KIS 436

RESULT 4
US-10-425-114-58982
; Sequence 58982, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong

APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E
APPLICANT: Tabaska, Jack E
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53113)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 58982
LENGTH: 437
TYPE: PRT
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: 700238628_FLI.pep
US-10-425-114-58982

Query Match 45.0%; Score 913; DB 4; Length 437;
Best Local Similarity 44.94; Pred. No. 7.3e-62;
Matches 190; Conservative 75; Mismatches 126; Indels 32; Gaps 7;

Qy 1 MGKSDFLSPKAIANRIKSGLOKLRWYCMQCKQCRDENGFKCHCMSESHORQLLASEN 60
Db 14 MGKHEFLTPKAIANRIKAKGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQVFGMA 73
Qy 61 PQQFMDYFSEEFNRNDFLELLRRRFGTKRVHNNIVYNEYISHREHIHNNATOWETLTDTFK 120
Db 74 PDVVEGFSEEFLESFLSLIRRAHRSRVAATVYNEYIADRHVHNMSTRWATLTFFVK 133
Qy 121 WLREGCLKVDPTPKGWIYQYIDRDPETIRROLELEKKKKQDLDDEKTAFTIEEQVRR- 179
Db 134 FLGREGYCKVEDTPKGMFTYIDRDSEQAVKRLKRIKSDMADDERQERMIARQIERA 193
Qy 180 -----GLEKQETPVPFTLSRENEEEKVTFLNKGAGGSAG--ATTSSSSSLGP 227
Db 194 HKSIAKPNNGGGAEGEPESGEYSDDDEQEPEDDSKEADKATGKIAIALQKAVPGP 253
Qy 228 SA-----LKL-----LGSASGGRK-----ESSQSSAQPAPKKKKSALDEIMELEEE-K 269
Db 254 KVNPFDDKPKMKFGFDEEDSGTRDQEKNELTKMGKDKVKAABAKRSALDELAKKEBEKAK 313
Qy 270 KRTARTDAMLQPGIVVVKIITKLGK-YHKKKGVVKEVIDRYTAVVQMTDSGRLKLDQT 328
Db 314 ERSNRKDYWLCPGIVVVKMSKSLAEKGYKQGVVKKVIDKYVGEIEMLESKHVLRVDQD 373
Qy 329 HLETVIPAGKRVLVNGGYRGNEGTLSEINEKAFSATIVETGPKLGRVVEGIOVEDIS 388
Db 374 ELETVPIQIGLVRVNGAYRGSNARLLSVDTKEFCAKQVKEKGLYDGKVLRAVEYEDIC 433
Qy 389 KLA 391
Db 434 KIS 436

RESULT 5
US-10-425-115-341788
Sequence 341788, Application US/10425115
Publication No. US20040214272A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants
FILE REFERENCE: 38-21(53222)B
CURRENT APPLICATION NUMBER: US/10/425,115
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 369326
SEQ ID NO 341788
LENGTH: 424

TYPE: PRT
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: MRT4577_74879C.1.pep
US-10-425-115-341788

Query Match 44.8%; Score 908; DB 4; Length 424;
Best Local Similarity 44.9%; Pred. No. 1.7e-61;
Matches 190; Conservative 73; Mismatches 128; Indels 32; Gaps 7;

Qy 1 MGKSDFLSPKAIANRIKSGLOKLRWYCMQCKQCRDENGFKCHCMSESHORQLLASEN 60
Db 1 MGKHEFLTPKAIANRIKAKGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQVFGMA 60
Qy 61 PQQFMDYFSEEFNRNDFLELLRRRFGTKRVHNNIVYNEYISHREHIHNNATOWETLTDTFK 120
Db 61 PDVVEGFSEEFLESFLSLIRRAHRSRVAATVYNEYIADRHVHNMSTRWATLTFFVK 120
Qy 121 WLREGCLKVDPTPKGWIYQYIDRDPETIRROLELEKKKKQDLDDEKTAFTIEEQVRR- 179
Db 121 LLGREGYCKVEDTPKGMFTYIDRDSEQAVKRLKRIKSDMADDERQERMIARQIERA 180
Qy 180 -----GLEKQETPVPFTLSRENEEEKVTFLNKGAGGSAG--ATTSSSSSLGP 227
Db 181 HKSIAKPNNGGGAEGEPESGEYSDDDEQEPEDDSKEADKATGKIAIALQKAVPGP 240
Qy 228 SA-----LKL-----LGSASGGRK-----ESSQSSAQPAPKKKKSALDEIMELEEE-K 269
Db 241 KVNPFDDKPKMKFGFDEEDSGTRDQEKNELTKMGKDKVKAABAKRSALDELAKKEBEKAK 300
Qy 270 KRTARTDAMLQPGIVVVKIITKLGK-YHKKKGVVKEVIDRYTAVVQMTDSGRLKLDQT 328
Db 301 ERSNRKDYWLCPGIVVVKMSKSLAEKGYKQGVVKKVIDKYVGEIEMLESKHVLRVDQD 360
Qy 329 HLETVIPAGKRVLVNGGYRGNEGTLSEINEKAFSATIVETGPKLGRVVEGIOVEDIS 388
Db 361 ELETVPIQIGLVRVNGAYRGSNARLLSVDTKEFCAKQVKEKGLYDGKVLRAVEYEDIC 420
Qy 389 KLA 391
Db 421 KIS 423

RESULT 6

US-10-739-930-8499
Sequence 8499, Application US/10739930
Publication No. US20040216190A1
GENERAL INFORMATION:
APPLICANT: Kovalic, David K.
TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
FILE REFERENCE: 38-21(53377)B
CURRENT APPLICATION NUMBER: US/10/739,930
CURRENT FILING DATE: 2003-12-18
NUMBER OF SEQ ID NOS: 11088
SEQ ID NO 8499
LENGTH: 423
TYPE: PRT
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: ZEAMA-23APR03-C7583_2.p
US-10-739-930-8499

Query Match 44.5%; Score 901.5; DB 5; Length 423;
Best Local Similarity 44.5%; Pred. No. 5.4e-61;
Matches 189; Conservative 72; Mismatches 127; Indels 37; Gaps 8;

Qy 1 MGKSDFLSPKAIANRIKSGLOKLRWYCMQCKQCRDENGFKCHCMSESHORQLLASEN 60
Db 1 MGKHEFLTPKAIANRIKAKGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQVFGMA 60
Qy 61 PQQFMDYFSEEFNRNDFLELLRRRFGTKRVHNNIVYNEYISHREHIHNNATOWETLTDTFK 120
Db 61 PDVVEGFSEEFLESFLSLIRRAHRSRVAATVYNEYIADRHVHNMSTRWATLTFFVK 120

```
Db 61 PDRVVEGSEFELESLIRRAHRSVAATVYVNEYIADRRHHVHMSTRWATLTFVK 120
QY 121 WLGREGLCVDETPKGYIYQIDRDETIRROLELEKKKKQDLDDEETAKFIESQVRG 180
Db 121 LLGREGYCKVEDTPKGFMTYIDRSEQAVKORLKRRIKSDMAEDERQERMIARQIERA 180
QY 181 LE-----GKEQETPVFTLESENHEE-EKVTFNLNKKGAGGSAGA--TTS 220
Db 181 HKSLAKPNDGAAGEPESGGSEE---YSGDDGSEEDVSKAEANKATGKTAIALQTAV 237
QY 221 KSSSLGPSALK-----LLGSAAGSKKE-----SSQSSAPAKKKKALDEIMELEB- 267
Db 238 RQKVNPFDEKVEKFRPEEDGLTGGEKDELAKKKGDKVKAADARRSALLDELKEEM 297
QY 268 EKRTARTDAWLOPGIVVKIITKLGKEK-YHKKGGVVKEVIDRYTAVVQVTDGDRCLKD 326
Db 298 AXERSNRKDYWLCPGIVVVKMSKLAEGYKQGVVRKVMKDYKVGIEIMLESKHVLRVD 357
QY 327 QTHLETVIPAPKRVVLVNGYRGNEGTLSEINEKAFSATIVETGPKLGRVRVEGIQYED 386
Db 358 QDELETVIPQIGGLVRILNGAYRGSNARLLSVDTKEFCAKVQVKEGLYDGVLRVAVKYBD 417
QY 387 ISKLA 391
Db 418 ICKIS 422

RESULT 7
US-10-424-599-168326
; Sequence 168326, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 168326
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_123013C.1.pep
US-10-424-599-168326

Query Match 30.2%; Score 613; DB 4; Length 176;
Best Local Similarity 64.3%; Pred. No. 3.1e-39;
Matches 108; Conservative 26; Mismatches 34; Indels 0; Gaps 0;

QY 1 MGSDFLSPKAIANRIKSGLOKLRYWCQKQCRDENGFKCHCMSESHORQLLASSEN 60
Db 1 MGNFELTPKAIANRIKAKGLQKLRYWCQKQCRDENGFKCHCMSEGHQRMQIFQGN 60
QY 61 PQQFMDYFSEFRNDFLELLRRRFGPKRVHNNIVYNEYISHREHIHMNATOWETLTDFTK 120
Db 61 PHRIVEGYSEESTPLEHMKSHRSRVAATVYVNEYINDRRHHIMNSTQWATLTFVK 120
QY 121 WLGREGLCVDETPKGYIYQIDRDETIRROLELEKKKKQDLDDEEK 168
Db 121 YLGRTGCKVEETPKGFMTYIDRDETITFKERMKNKRIKADWDVEEK 168

RESULT 8
US-10-437-963-191916
; Sequence 191916, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.

Db 61 PDRVVEGSEFELESLIRRAHRSVAATVYVNEYIADRRHHVHMSTRWATLTFVK 120
QY 121 WLGREGLCVDETPKGYIYQIDRDETIRROLELEKKKKQDLDDEEK 180
Db 121 LLGREGYCKVEDTPKGFMTYIDRSEQAVKORLKRRIKSDMAEDERQERMIARQIERA 180
QY 181 LE-----GKEQETPVFTLESENHEE-EKVTFNLNKKGAGGSAGA--TTS 220
Db 181 HKSLAKPNDGAAGEPESGGSEE---YSGDDGSEEDVSKAEANKATGKTAIALQTAV 237
QY 221 KSSSLGPSALK-----LLGSAAGSKKE-----SSQSSAPAKKKKALDEIMELEB- 267
Db 238 RQKVNPFDEKVEKFRPEEDGLTGGEKDELAKKKGDKVKAADARRSALLDELKEEM 297
QY 268 EKRTARTDAWLOPGIVVKIITKLGKEK-YHKKGGVVKEVIDRYTAVVQVTDGDRCLKD 326
Db 298 AXERSNRKDYWLCPGIVVVKMSKLAEGYKQGVVRKVMKDYKVGIEIMLESKHVLRVD 357
QY 327 QTHLETVIPAPKRVVLVNGYRGNEGTLSEINEKAFSATIVETGPKLGRVRVEGIQYED 386
Db 358 QDELETVIPQIGGLVRILNGAYRGSNARLLSVDTKEFCAKVQVKEGLYDGVLRVAVKYBD 417
QY 387 ISKLA 391
Db 418 ICKIS 422

RESULT 9
US-10-767-701-57151
; Sequence 57151, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 57151
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 30963818.pep
US-10-767-701-57151

Query Match 22.1%; Score 449; DB 4; Length 120;
Best Local Similarity 66.7%; Pred. No. 8.4e-27;
Matches 80; Conservative 15; Mismatches 25; Indels 0; Gaps 0;

QY 1 MGSDFLSPKAIANRIKSGLOKLRYWCQKQCRDENGFKCHCMSESHORQLLASSEN 60
Db 1 MGNFELTPKAIANRIKAKGLQKLRYWCQKQCRDENGFKCHCMSEGHQRMQVFGMA 60
QY 61 PQQFMDYFSEFRNDFLELLRRRFGPKRVHNNIVYNEYISHREHIHMNATOWETLTDFTK 120
Db 61 PDRVVEGSEFELESLIRRAHRSVAATVYVNEYIADRRHHVHMSTRWATLTFVK 120
```

```
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 191916
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_88190C.1.pep
US-10-437-963-191916

Query Match 26.2%; Score 531; DB 4; Length 138;
Best Local Similarity 66.7%; Pred. No. 4.7e-33;
Matches 92; Conservative 22; Mismatches 24; Indels 0; Gaps 0;

QY 1 MGSDFLSPKAIANRIKSGLOKLRYWCQKQCRDENGFKCHCMSESHORQLLASSEN 60
Db 1 MGNFELTPKAIANRIKAKGLQKLRYWCQKQCRDENGFKCHCMSEGHQRMQVFGQA 60
QY 61 PQQFMDYFSEFRNDFLELLRRRFGPKRVHNNIVYNEYISHREHIHMNATOWETLTDFTK 120
Db 61 PDRVVEGSEFELESLIRRAHRSVAATVYVNEYIADRRHHVHMSTRWATLTFVK 120
QY 121 WLGREGLCVDETPKGY 138
Db 121 FLGREGHCKVEDTPKGF 138

RESULT 9
US-10-767-701-57151
; Sequence 57151, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 57151
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 30963818.pep
US-10-767-701-57151

Query Match 22.1%; Score 449; DB 4; Length 120;
Best Local Similarity 66.7%; Pred. No. 8.4e-27;
Matches 80; Conservative 15; Mismatches 25; Indels 0; Gaps 0;

QY 1 MGSDFLSPKAIANRIKSGLOKLRYWCQKQCRDENGFKCHCMSESHORQLLASSEN 60
Db 1 MGNFELTPKAIANRIKAKGLQKLRYWCQKQCRDENGFKCHCMSEGHQRMQVFGMA 60
QY 61 PQQFMDYFSEFRNDFLELLRRRFGPKRVHNNIVYNEYISHREHIHMNATOWETLTDFTK 120
Db 61 PDRVVEGSEFELESLIRRAHRSVAATVYVNEYIADRRHHVHMSTRWATLTFVK 120
```



```

RESULT 10
US-10-106-698-5464
; Sequence 5464, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5464
; LENGTH: 84
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-106-698-5464

Query Match      17.8%; Score 361; DB 4; Length 84;
Best Local Similarity 94.6%; Pred. No. 3.2e-20;
Matches 70; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 318 DSGDKLQDTHLETVPAPGKRVLVNGGYRGNEGTLSEINAKFSATIVETGPKLGR 377
Db 11 ESGDKLQDTHLETVPAPGKRVLVNGGYRGNEGTLSEINAKFSATIVETGPKLGR 70

QY 378 RVEGIQVEDISKLA 391
Db 71 RVEGIQVEDISKLA 84

RESULT 11
US-10-424-599-196166
; Sequence 196166, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 196166
; LENGTH: 162
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_19163C.1.pep
; US-10-424-599-196166

Query Match      16.0%; Score 324; DB 4; Length 162;
Best Local Similarity 46.8%; Pred. No. 5.5e-17;
Matches 72; Conservative 27; Mismatches 45; Indels 10; Gaps 3;

QY 240 KRKESQSSAPAKKKSALDEIMLEBEEKK-RTARTDAMIQPGIVVKKIITKKLGEK-YH 297
Db 17 KRKESGGG-----KSALEDMMREBEEKKKEKINRKYDLWHEGIVVVKMSKVLAEKGY 68

QY 298 KKKGVKVEIDRYTAVVMTDSGRDLKLDQTHLETVPAPGKRVLVNGGYRGNEGTLSE 357
Db 69 KQGVKRVKIDKYVGEIEMLESKHLVLDQAELETVPQVGGRVKIVNGAYRGSIAKLGG 128

QY 358 INEKAFSATIVETGPKLGRVEGIQVEDISKLA 391

```

```

Db 129 VDTDNFCAKQVIEKGAYDGRVLKSVYEYEDICKVA 162

RESULT 12
US-10-425-115-300949
; Sequence 300949, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 300949
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(123)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_37535C.1.pep
; US-10-425-115-300949

Query Match      11.8%; Score 239; DB 4; Length 123;
Best Local Similarity 46.5%; Pred. No. 1.4e-10;
Matches 46; Conservative 14; Mismatches 39; Indels 0; Gaps 0;

QY 10 KAIANRIKSGKGLRWYCOMCQKCRDENGFKCHCMSSSHQQLLIASENPQOFMDYFS 69
Db 25 KAIANRIKSGKGLRWYCOMCQKCRDENGFKCHCMSSSHQQLLIASENPQOFMDYFS 84

QY 70 EFRNDFLELLRRRFGTKRVHNNIVNYEYISHREHIMN 108
Db 85 KEFLXSFLSFIRHADRHSRLTATXVYNYKYIVDRYHVMN 123

RESULT 13
US-10-425-115-230934
; Sequence 230934, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 230934
; LENGTH: 107
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_14220SC.1.pep
; US-10-425-115-230934

Query Match      10.9%; Score 221.5; DB 4; Length 107;
Best Local Similarity 40.6%; Pred. No. 2.6e-09;
Matches 43; Conservative 28; Mismatches 34; Indels 1; Gaps 1;

QY 287 IITKGLGEK-YHKKGKGVKVEIDRYTAVVMTDSGRDLKLDQTHLETVPAPGKRVLVN 345

```

```

Db      1  VMKSLAEKYYKQGVVRKMDKYVEIMLESKHVLRVQDELETVPQIGLLRIILN 60

Qy      346  GGYRGNEGTLSESINEKAFSATTIVETGPLKGRVEGIQYVEDISKLA 391
           |||:  ||::  ||::  ||::  ||::  ||::  ||::  ||::  ||::
Db      61  GAYRGSNARLLSVDTTEKFCAKVQEKGYDGKVLRAVEYEDICKIS 106

RESULT 14
US-10-424-599-196168
; Sequence 196168, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy, Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)/B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 196168
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(109)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_19165C.1.pep
US-10-424-599-196168

```

	Query Match	10.3%;	Score 209;	DB 4;	Length 109;
	Best Local Similarity	41.3%;	Pred. No. 2.4e-08;		
	Matches	45;	Conservative 23;	Mismatches 41;	Indels 0; Gaps 0;
Qy	283	I V K I I T K L G E Y H K K G V V K E V I D R Y T A V V K W T M S G D R L K L Q O T H L E T V I P A G K R V L	342		
		: : : : : :			
Dd	1	I V K V M N K L K G S H N Q K G V R K V I V K Y V G E I E M L E S K H V L R V D Q A K L E T V I P O V G G R V K	60		
		: : : : : :			
Qy	343	V L N G C Y R G N E G T L E S I N E K A F S A T I V I E T G P L K G R R V E G T Q Y E D I S K L A	391		
		: : : : : :			
Dd	61	I V A N G A Y R G S I A K L G V D T N N F C A K G H I E K Y V D G R V L K A V E Y E A L C K V A	109		
		: : : :			

```

RESULT 15
US-10-425-114-46572
; Sequence 46572, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(531313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 46572
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700442544_FLI.pep
US-10-425-114-46572
Query Match 7.3%; Score 149; DB 4; Length 278;

```

```

Best Local Similarity 35.0%, Pred. No. 0.0036;
Matches 49; Conservative 20; Mismatches 51; Indels 20; Gaps 5;

Qy 238 SGRKQSSQAQPAKKKSGALDEIMLE-EEKKRTAR-----TDA-----WLQPG 282
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109 SRDMEEDSSAQ----KKRYNEERSEKEGKRVNRGSGRECTSNASDTHSNVWLQSH 164
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 283 IVVKIIFPKLGEKHHKKGVVKEVIDRYTAVVKNMTSGDRLK-LDQTHLETVIPAPGKRV 341
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 165 IRVRIVSQKLSKRLYLKMGKGVVDVVGPTKCDIMWDGSELVQGVQDMLETPLPTNGRV 224
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 342 LVNLGGYRGNEGTLTSESINEK 361
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 225 LVLYGRKHGGLYGHVLEKNSK 244
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: November 23, 2005, 16:39:34
Job time : 115.709 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

QM protein - protein search, using sw model

Run on: November 23, 2005, 16:28:34 ; Search time 4.98725 Seconds
(without alignments)
237.628 Million cell updates/sec

Title: US-09-555-529-25
Perfect score: 2028
Sequence: 1 MGKSDFLSPKAIANRIKSG.....GPKGRVRGVIQVEDISKLA 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 17545 seqs, 3030971 residues

Total number of hits satisfying chosen parameters: 17545

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA New:
1: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pdb:
2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pdb:
3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pdb:
4: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pdb:
5: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pdb:
6: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pdb:
7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pdb:
8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pdb:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	110.5	5.4	853	1	US-10-821-234-1110
2	105	5.2	885	1	US-10-793-626-1660
3	104	5.1	676	7	US-11-135-855-28
4	104	5.1	709	7	US-11-074-176-158
5	104	5.1	717	7	US-11-135-855-29
6	100	4.9	672	1	US-10-689-742-70
7	99.5	4.9	1448	1	US-10-485-517-212
8	94.5	4.7	583	1	US-10-793-626-1358
9	93.5	4.6	756	7	US-11-074-176-202
10	93.5	4.6	1189	7	US-11-074-176-134
11	93.5	4.6	1279	1	US-10-793-626-3188
12	93	4.6	989	1	US-10-821-234-975
13	92.5	4.6	752	1	US-10-793-626-348
14	91	4.5	677	1	US-10-982-545-12
15	90.5	4.5	618	1	US-10-793-626-860
16	89.5	4.4	422	1	US-10-821-234-1313
17	88.5	4.4	296	1	US-10-131-826A-26
18	87.5	4.3	1618	1	US-10-984-645-2
19	87	4.3	416	1	US-10-793-626-2
20	87	4.3	1437	7	US-11-074-176-96
21	87	4.3	5024	1	US-10-793-626-2964
22	86.5	4.3	594	1	US-10-131-826A-10
23	86.5	4.3	594	1	US-10-485-517-416
24	86.5	4.3	1299	1	US-10-821-234-1145
25	86.5	4.3	2897	1	US-10-499-715-2

26	85	4.2	419	1	US-10-793-626-2516	Sequence 2516, Ap
27	85	4.2	745	1	US-10-793-626-1500	Sequence 1500, Ap
28	84.5	4.2	309	7	US-11-074-176-276	Sequence 276, App
29	84	4.1	431	7	US-11-074-176-132	Sequence 132, App
30	84	4.1	770	1	US-10-982-545-15	Sequence 15, Appl
31	84	4.1	770	1	US-10-789-273-38	Sequence 38, Appl
32	83.5	4.1	654	7	US-11-046-668-4	Sequence 4, Appl
33	83.5	4.1	1614	1	US-10-821-234-903	Sequence 903, App
34	83	4.1	930	1	US-10-821-234-1188	Sequence 1188, Ap
35	82.5	4.1	691	1	US-10-131-826A-16	Sequence 16, Appl
36	82.5	4.1	703	1	US-10-821-234-963	Sequence 963, App
37	81.5	4.0	231	1	US-10-957-569-21	Sequence 21, Appl
38	81	4.0	196	1	US-10-793-626-2088	Sequence 2088, Ap
39	80.5	4.0	752	1	US-10-793-626-1138	Sequence 1138, Ap
40	80.5	4.0	874	7	US-11-012-762-8	Sequence 8, Appl
41	80.5	4.0	974	1	US-10-821-234-1152	Sequence 1152, Ap
42	80	3.9	469	1	US-10-821-234-1151	Sequence 1151, Ap
43	80	3.9	615	1	US-10-982-545-14	Sequence 14, Appl
44	80	3.9	674	1	US-10-507-275-9	Sequence 9, Appl
45	80	3.9	683	7	US-11-046-668-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-10-821-234-1110
; Sequence 1110, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; PRIOR FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 1110
; LENGTH: 853
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-821-234-1110

Query Match	5.4%	Score 110.5;	DB 1;	Length 853;
Best Local Similarity	28.3%	Pred. No. 0.26;		
Matches	58;	Conservative	24;	Mismatches 58; Indels 65; Gaps 11;
Qy	120	KWLGREGLC-KVDETPKGWIQYIDRDPETIRQLELEKK-----KK	160	
Db	561	KALWFQRCVKVLSBK--YKLVLRIP---NRGIDLLKKDKRKRSPDGKESPSDKK	615	
Qy	161	QDLDDDEKTKAFI---EEQVRGLEGK-----EQETPVFTLSRE---NEBEKVTFN	206	
Db	616	SKTDGSKQTESSTEGKEGEGEDGKOTKDDQTEQEPNNLLESDELLVDEEE-----	670	
Qy	207	LNKGAGSGAGATTSKSSLGPSA-LKLLGSAAGSKRKESQ-----SSAQPAKKKKA	258	
Db	671	-----AAALLSGSSVGDETDLNLDVADGDKPEPSDKAVKDGSAASAAKKLKK	722	
Qy	259	LDEIMELEBEKKRTARTDQWLQPGI	283	
Db	723	VDKIEELDQB-----NEAALENGI	741	

RESULT 2
US-10-793-626-1660
; Sequence 1660, Application US/10793626
; Publication No. US20050255478A1

```

; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1660
; LENGTH: 885
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-1660

Query Match          5.2%; Score 105; DB 1; Length 885;
Best Local Similarity 17.9%; Pred. No. 0.7;
Matches 67; Conservative 76; Mismatches 136; Indels 96; Gaps 15;

QY 48 ESHOROLLASENPQPFMDYFSEFRNDFLELLRRRFGTKRVHNNIVYNEYISHRE--- 103
DB 486 EDYQRO---TAPERFQF-----KEQKTQAEQLSEQYFSASQQYNNJ---KEQVHHHELELD 535
QY 104 -----HIHMNATOWETLDTFTKWLREGCLCKVDETPKGWYIQYIDRDPETIRRO--- 152
DB 536 RLKQTEAHLKNEHEFE---PEKNDGYQS-----DKSKETLKEKQNH 575
QY 153 LELEKKKKQDLDEETAKFIEB-----QVRGLGKQEQETPVFTELSRNEEEKVTYN 206
DB 576 IEIQOQLKQLESDIERYTQLSKEGKASTHTQTQOQLHQKQSDLAVVKR--RIKSQKVYER 633
QY 207 LNKGAGGSAGATTSSKSSSLGP-SALKLGSAAAGKRGKSSQ-----SSAQPAKKK 255
DB 634 LKQLSDSERQKTEVNEKIKLFNSDEMGMKDAPEKLEIQOQENVRONLNQOLSEIKQ 693
QY 256 KSAALDIMELEEBEKKRTARTDAWLQPGVIVVVIITKKLGGKYHKKGKGVKVEDIDRYTAVVK 315
DB 694 RDLNKEKIEINSQLKCHQDI-----LSIENHYQDIKAKQSKLDVLINHAIDLNDTYQ 748
QY 316 MTDGDRKLQDLTHLETVPAPGRVVLVNGVGRVNEGTLESINEKAFSATIVI-ETGPL 374
DB 749 LTVARAMEYD-----SDETIDNLNRKKVKLTGKTMTIDELGPV 784
QY 375 KGRRVGIEYEDISK 389
DB 785 NLNAIE--QFEELNE 797

RESULT 3
US-11-135-855-28
; Sequence 28, Application US/11/135855
; Publication No. US2005025557A1
; GENERAL INFORMATION:
; APPLICANT: SMITHKLINE BEECHAM CORPORATION
; APPLICANT: SMITHKLINE BEECHAM P.L.C.
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50013
; CURRENT APPLICATION NUMBER: US/11/135,855
; CURRENT FILING DATE: 2005-05-24
; PRIOR APPLICATION NUMBER: US/10/203,708
; PRIOR FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: PCT/US01/04703
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/182,172
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: 60/186,084
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 28
; LENGTH: 676
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-135-855-28

Query Match          5.1%; Score 104; DB 7; Length 676;
Best Local Similarity 21.3%; Pred. No. 0.59;
Matches 61; Conservative 46; Mismatches 89; Indels 90; Gaps 12;

QY 144 RDPETIRRQLELEKKKQDLDEETAKFIEBQVRRGLEKQEQETPVFTELSRNEEEKV 203
DB 318 RDEARRELEARRREQ-----BEELRRLEQSKKEKERRERRADRGEAE-- 363
QY 204 TPNLNKGAGGSAGATTSSKSSSLGPSALKLGSAAAGKRGKSSQSAQAQAKKKKSDALDEIM 263
DB 364 -----RSGSGSGDELREDD---EPVKRGRKGRGRGPPSSSDSEPEA----- 403
QY 264 ELEEEKKETAR-----TDAMLQPGIVVKII-----TKKLGEKY----- 296
DB 404 ELEREAKKSAKXQSSSTEPARKPGQKEKRVPEEKQQAQKPVKVRTRRRSEGFMSMDRKV 463
QY 297 -HKKGVVKEVIDRYTAVVGMTDSGDRKLQDLTHLETVIPA-----PGKRVLVNLGG 347
DB 464 EKKKEPSVEEKLQKLHSEIKFA----LKVDSPOVKRCLNALEELGLTQLVTSQILQKNTD 518
QY 348 -----VRGNEGTLESINEKAFSATIVIETGPKLGRRVGEGIQ 383
DB 519 VVATLKLIRRYKAN----KDVMEKAAEYVYTRLKSRLV-GPKIEAVQ 559

RESULT 4
US-11-074-176-158
; Sequence 158, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
; APPLICANT: Klaenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Peril, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; TITLE OF INVENTION: Stress-Related Proteins and Uses Therefore
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074,176
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: 60/551,161
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 158
; LENGTH: 709
; TYPE: PRT
; ORGANISM: Lactobacillus acidophilus
US-11-074-176-158

Query Match          5.1%; Score 104; DB 7; Length 709;
Best Local Similarity 20.4%; Pred. No. 0.63;
Matches 83; Conservative 49; Mismatches 127; Indels 148; Gaps 17;

QY 65 MDYFSEFRNDPFLQL-----LRRRFCTKEVHNNIVYNEIHSRHHMNAQWETLTDPTK 120
DB 1 MAYFNDFDNLNLFNLSNFFNDDFGSRSDNN----- 32
QY 121 WLGREGLCKVDETPKGWYIQYIDRDPETIRR--QLELEKKKKQDLDDBEKTAKF-----I 173
DB 33 --GNSGSI PMN-----YSSNNSSSPRTIQNPQPNKPIGVDLVERAKNNKFPDPIGR 84
QY 174 EQQVRGLE--GKEQETPVF-----TELSENEEEK-----VTFNLN 208
DB 85 DEQIDNVIEILSRKKNNPNVLIGPAGVGKTSIVEGLAERIAAGNVPKAMNHHISVNI 144
QY 209 KGAGGS-----AGATTSSKSSSLGPSALKL 233
```

```

; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1660
; LENGTH: 885
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-1660

Query Match          5.1%; Score 104; DB 7; Length 676;
Best Local Similarity 21.3%; Pred. No. 0.59;
Matches 61; Conservative 46; Mismatches 89; Indels 90; Gaps 12;

QY 144 RDPETIRRQLELEKKKQDLDEETAKFIEBQVRRGLEKQEQETPVFTELSRNEEEKV 203
DB 318 RDEARRELEARRREQ-----BEELRRLEQSKKEKERRERRADRGEAE-- 363
QY 204 TPNLNKGAGGSAGATTSSKSSSLGPSALKLGSAAAGKRGKSSQSAQAQAKKKKSDALDEIM 263
DB 364 -----RSGSGSGDELREDD---EPVKRGRKGRGRGPPSSSDSEPEA----- 403
QY 264 ELEEEKKETAR-----TDAMLQPGIVVKII-----TKKLGEKY----- 296
DB 404 ELEREAKKSAKXQSSSTEPARKPGQKEKRVPEEKQQAQKPVKVRTRRRSEGFMSMDRKV 463
QY 297 -HKKGVVKEVIDRYTAVVGMTDSGDRKLQDLTHLETVIPA-----PGKRVLVNLGG 347
DB 464 EKKKEPSVEEKLQKLHSEIKFA----LKVDSPOVKRCLNALEELGLTQLVTSQILQKNTD 518
QY 348 -----VRGNEGTLESINEKAFSATIVIETGPKLGRRVGEGIQ 383
DB 519 VVATLKLIRRYKAN----KDVMEKAAEYVYTRLKSRLV-GPKIEAVQ 559

RESULT 4
US-11-074-176-158
; Sequence 158, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
; APPLICANT: Klaenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Peril, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; TITLE OF INVENTION: Stress-Related Proteins and Uses Therefore
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074,176
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: 60/551,161
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 158
; LENGTH: 709
; TYPE: PRT
; ORGANISM: Lactobacillus acidophilus
US-11-074-176-158

Query Match          5.1%; Score 104; DB 7; Length 709;
Best Local Similarity 20.4%; Pred. No. 0.63;
Matches 83; Conservative 49; Mismatches 127; Indels 148; Gaps 17;

QY 65 MDYFSEFRNDPFLQL-----LRRRFCTKEVHNNIVYNEIHSRHHMNAQWETLTDPTK 120
DB 1 MAYFNDFDNLNLFNLSNFFNDDFGSRSDNN----- 32
QY 121 WLGREGLCKVDETPKGWYIQYIDRDPETIRR--QLELEKKKKQDLDDBEKTAKF-----I 173
DB 33 --GNSGSI PMN-----YSSNNSSSPRTIQNPQPNKPIGVDLVERAKNNKFPDPIGR 84
QY 174 EQQVRGLE--GKEQETPVF-----TELSENEEEK-----VTFNLN 208
DB 85 DEQIDNVIEILSRKKNNPNVLIGPAGVGKTSIVEGLAERIAAGNVPKAMNHHISVNI 144
QY 209 KGAGGS-----AGATTSSKSSSLGPSALKL 233
```

Db 145 DMVAGSLRGSFPERLKKVIDKAKSDPNIVLFDLHNIVGAGSTSDENN-GDAANILK 203
Qy 234 GSAAGKRKSSOSSAQAOKKXKALDEIMELBEEKRTARTDANLQ-----GIVVKII- 288
Db 204 PALASGELKIJGATTT-----SEFORIEKDPALSRFQAVQVPEPSTDAIKILE 253
Qy 289 -TKKLGKGYHKKGWKEVIDRYTAVVKMTSDGRLKLDQTHLETVIPA---PGRVLVL 344
Db 254 GLKKYEDYH-----VKYTD--DSLKLAVELSERYLOGRYLPDKAIDLM 296
Qy 345 NGYGRNGEGTLESINEKAFSATIVETGPLKGRVREGIQYEDISKLA 391
Db 297 DEAGAKKALLVQPTDEKSLKNQI-----SALEAKKAAEAAEDYDKAA 339

RESULT 5
US-11-135-855-29
; Sequence 29, Application US/11135855
; Publication No. US2005025557A1
; GENERAL INFORMATION:
; APPLICANT: SMITHKLINE BEECHAM CORPORATION
; APPLICANT: SMITHKLINE BEECHAM P.L.C.
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50013
; CURRENT APPLICATION NUMBER: US/11/135,855
; CURRENT FILING DATE: 2005-05-24
; PRIOR APPLICATION NUMBER: US/10/203,708
; PRIOR FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: PCT/US01/04703
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/182,172
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: 60/186,084
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 29
; LENGTH: 717
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-135-855-29

Query Match 5.1%; Score 104; DB 7; Length 717;
Best Local Similarity 21.3%; Pred. No. 0.63;
Matches 61; Conservative 46; Mismatches 89; Indels 90; Gaps 12;
Qy 144 RDPETIRROLEKKKKQDLDDDEEKTAKFIEEQVRRGLEKGEQETPVFTLSRENEBEKV 203
Db 359 RRDARRLEARRREQ-----BEELRLREQEKEKERRERADRGAE-- 404
Qy 204 TFLNKGAGSGAGATTSSKSLGPSALKLIGSAAAGKRKSSQAQAPAKKKKXALDEIM 263
Db 405 -----RSGSGSGDELREDD---EPVKRGRKGRGPPSSSDSEPEA----- 444
Qy 264 ELEBEKGTAR-----TDWLQPGIVVKII-----TKKLGKY----- 296
Db 445 ELEREAKKSAKPOSSSTEPARKQKQKVRPEEKQAKPVKVERTKRSEGSMDRKV 504
Qy 297 -HKKGWKEVIDRYTAVVKMTSDGRLKLDQTHLETVIPA-----PGRVLVLNGG 347
Db 505 EKKKEPSVEEKLLQKLSIEKFA-----LKVSDPVKRCINALEELGTQVTSQILQKNTD 559
Qy 348 -----YRGNBGTLESINEKAFSATIVETGPLKGRVREGIQ 383
Db 560 VVATLKKIRRYKAN-----KDVMEKAAEVYTRLKSRLV-GPKIEAVQ 600

RESULT 6
US-10-689-742-70
; Sequence 70, Application US/10689742
; Publication No. US20050250180A1
; GENERAL INFORMATION:

; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M
; APPLICANT: LaVallie, Edward R
; APPLICANT: Racie, Lisa A
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 00766.000091.10
; CURRENT APPLICATION NUMBER: US/10/689,742
; CURRENT FILING DATE: 2003-10-22
; PRIOR APPLICATION NUMBER: 09/746,783
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 70
; LENGTH: 672
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (107)..(107)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (111)..(111)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (117)..(118)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (645)..(645)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-10-689-742-70

Query Match 4.9%; Score 100; DB 1; Length 672;
Best Local Similarity 23.6%; Pred. No. 1.1;
Matches 70; Conservative 41; Mismatches 107; Indels 78; Gaps 14;
Qy 33 KQCRDENGPKCHWSESHORQLLASENPOQFMDYFSEFPRNDFLELLRRRFGTKRVHNN 92
Db 424 QQCAEEBG-----DIEWQLNASVLVAQAQVKLRDF--ESAVNNFEKALER---AKLVHNN 473
Qy 93 IVNEYISHREHIHMNATQWETLTDFTYKWLGRGLCKVD-----ETPKGYIYQIDR-- 144
Db 474 EAQAIIIS-----ALDDANKGIIRE-LRKTNYVENLKEKSEGEASLYEDRII 519
Qy 145 -----DPEITIRROLE-LEKKKKQDLDDDEEKTAKFIEEQVRRGLEKGEQETPVFT 192
Db 520 TREKDMRVRDEPEKVVVKQMDHSEDEKETDEDDA-----FGEALQSPASGKQSV----- 569
Qy 193 ELSRENEBEKVTFNLNKGAGSGAGATTSSKSLGPSALKLIGSAAAGKRKSSQAQAPAKKKKX 252
Db 570 -----EAGKARSDLGAVAKGLSGELGTRG---GETGRKLLGAG---RRSREIYRRPS 616
Qy 253 KKKKXAL-----DEIMELBEEKRTARTDANLQPGIVVKIIITKKLGKGYHKK 299
Db 617 GELFQRLSGEFSRQEPPELKKLSEVGRRXPEELGKTQFGEIGE--TKKTGNMEMSKE 670

RESULT 7
US-10-485-517-212
; Sequence 212, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynex Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides


```
Best Local Similarity 17.4%; Pred. No. 6.2;
Matches 75; Conservative 84; Mismatches 145; Indels 128; Gaps 16;

QY 49 SHORQLLASNPQQFMDY-----FSEFRNDLELLRRFGTKRVHNNI 93
Db 510 SVQKQKGLVSPESMDNYSHPHELAVTBEEINVVLKQDLQNALSESRNEKVRLEBKL 569
QY 94 VYNE-----YISHREHHMNAQTWLTDFTKWL-GREGLCCKVDETP 134
Db 570 VEREKGTWIKPPVEEYEEKSSYSYVNNWK-----EKAFLPEKYQEAQBEIMKLKDTL 624
QY 135 KGWYIQYDRDPTIRQL-----ELEKKKKQLDDP-----SEKATAP 172
Db 625 KSMQTQASDEADMEKAMNMIDELNKQVSELSQLYKEAQAELEDYRKRSLESDVTARY 684
QY 173 IEQVRRGL-----EGKEQETPVFTFL-----SRENEEKVTFNLN 208
Db 685 IHKAEHEKLMQNTVSRKAEDALSEMKSQYSKVLNLTQLKQLVDAQKENSVSITEHLQ 744
QY 209 KGAGGSAGATTSSKSLGSPALKLGSAAAGKRSQSSQAQPAKKKSGALDEIMLEBEE 268
Db 745 V-----ITTLR-----TAAKEMBEKISNLKHLASKEVEVAKLEKQLLEE----- 784
QY 269 KKTARTDAW-----IQPGI--VVKIITKLGKGYHKKGVVKEVIDRYTAVVWMTD 318
Db 785 --KAAMTDMVPRSSYEKQLSSLESESVLASKLESVKEKEV-----HSEVVQIRS 835
QY 319 SGRDLKLDOTHLSTVPAPOKRVLVNGGYRGNEGTLSEINAKAFSATTVIETGPKLGRR 378
Db 836 EYSQVREKENIOTLKSKEQVNEQLQKFOQAQBELAEMKRYAESSSKLEED---KDKK 892
QY 379 VEGIOYEDISKL 390
Db 893 INEMS-KEVTKL 903

RESULT 13
US-10-793-626-348
; Sequence 348, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMBERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 348
; LENGTH: 752
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-348

Query Match 4.6%; Score 92.5; DB 1; Length 752;
Best Local Similarity 18.1%; Pred. No. 4.7;
Matches 62; Conservative 73; Mismatches 112; Indels 95; Gaps 16;

QY 123 GREGLCCKVDTPKGYI--QYIDRDP-ETIRRL-----ELEK-----KKQDLDDEEK 168
Db 174 GIEGIKAVETGKVVRSRVDEEPLRSRKGKLIITEIPYEVNKSLSLVKRIDELADKK 233
QY 169 TAKFIE---EQVRGLE-----GKEQETPVFTELSRNEEEKVTFNLNKGAGGSAGATT 220
Db 234 VDGIVEVRDRTDTRLATLAIELKDNANSEIKNLYKNLQISYFNFMVA----- 284
QY 221 KSSSLGFSALKLIG-----SASGKRKESQSSQAQPAKKKKSALDE 261
```

```
Db 285 -----ISEGRPKLMGLREILIESYLNHQIEVVTNRTRYDLEQAEKRMHIIVEGLMKALSILDE 340
QY 262 IMEL--EEBKRTAR-----TDAWLQPGIIVVKII-----TKKLGEKYHKKGVVK 304
Db 341 VITALIRSNKNKDAKDNLVAEYDFTFEAQAEALVQLVRLTWTDTIEALKKEHEELEALIK 400
QY 305 E---VIDRYTAV-----VKMTDSGDRL-----KLDQTHLETVPAPOKRVVLNG 346
Db 401 ELRNILDHEALLAVIKDELNEIKKFKVDRLSTTEAEISEIKIDKEVMVPSEVIL--- 457
QY 347 GYRGNEGTLSEINAKAFSATTVIETGPKLGRRVEGIOYEDIS 388
Db 458 -SLTQHGVIKTRTSRFSNAGVTEIGLKGDRL--LKHSVN 496

RESULT 14
US-10-982-545-12
; Sequence 12, Application US/10982545
; Publication No. US20050244890A1
; GENERAL INFORMATION:
; APPLICANT: Davies, Huw Alun
; APPLICANT: McGuire, James
; APPLICANT: Simonsen, Anja Hviid
; APPLICANT: Blennow, Kaj
; APPLICANT: Podust, Vladimir
; APPLICANT: Ciphergen Biosystems, Inc.
; TITLE OF INVENTION: Biomarkers for Alzheimer's Disease
; FILE REFERENCE: 016866-011550US
; CURRENT APPLICATION NUMBER: US/10/982,545
; CURRENT FILING DATE: 2004-11-06
; PRIOR APPLICATION NUMBER: US 60/518,360
; PRIOR FILING DATE: 2003-11-07
; PRIOR APPLICATION NUMBER: US 60/526,753
; PRIOR FILING DATE: 2003-12-02
; PRIOR APPLICATION NUMBER: US 60/546,423
; PRIOR FILING DATE: 2004-02-19
; PRIOR APPLICATION NUMBER: US 60/547,250
; PRIOR FILING DATE: 2004-02-23
; PRIOR APPLICATION NUMBER: US 60/558,896
; PRIOR FILING DATE: 2004-04-02
; PRIOR APPLICATION NUMBER: US 60/572,617
; PRIOR FILING DATE: 2004-05-18
; PRIOR APPLICATION NUMBER: US 60/586,503
; PRIOR FILING DATE: 2004-07-08
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 677
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Chromogranin B precursor
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)..(20)
; OTHER INFORMATION: signal peptide
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (21)..(677)
; OTHER INFORMATION: Chromogranin B (Secretogranin I)
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (326)..(385)
; OTHER INFORMATION: biomarker peptide 7258 Da, processed fragment of
; OTHER INFORMATION: Chromogranin B
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (341)
; OTHER INFORMATION: Xaa = sulfotyrosine
US-10-982-545-12

Query Match 4.5%; Score 91; DB 1; Length 677;
Best Local Similarity 22.6%; Pred. No. 5.3;
```


Matches 60; Conservative 33; Mismatches 119; Indels 54; Gaps 10;	
Qy 111 QWETLDTFTWLGREGUCKVDETPKGYIQYIDBDPTIRROLEKKKKQDLDDDEKTA 170	
Db 113 QGPTKADTEKWAEGGHSRERADPQWLSYFSDSQVSEEVKTRHSEKSDREDEEE--- 169	
Qy 171 KFTIEQVRRGLEG-----KEQETPFVTELSRENE-----EEKVTFNLNKGAGGSA 215	291 KLGEKYH-----KKKGVV 303
Db 170 ---GENYQKGERGDSSEKHLBEPGETQNAFLNERQASAIKKEELVARSETHAAGHSQ 226	584 LMGQKSEGMLLSAEKDGVL 602
Qy 216 GATTSKSSSLGPSALKLLGSAAGKRKESQSSAQPAKKKKKSAI-DEIMLEBEKKRT-- 272	
Db 227 EKTHREKSSQES-----GSEAGSQNHPOESKQPRSQBESBEGEEDATSEVDKRTRP 281	
Qy 273 -----ARTDAWLQPGIIVKLIITKLGEKYHKKGVVKEVIDRYTAVVKTDSGDRCLKD 326	
Db 282 RHHGRSRPDRSSQGSGLPS-----BEKGHPQE--ESESNNVSMASLGEKRDHH 328	
Qy 327 QTHLETVIPAF--GKRVLVNLGGYRG 350	
Db 329 STHYRASEEPEXGEEI-----KGYPG 350	
RESULT 15	
US-10-793-626-860	
; Sequence 860, Application US/10793626	
; Publication No. US20050255478A1	
; GENERAL INFORMATION:	
; APPLICANT: KIMMERLY, WILLIAM JOHN	
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS	
; FILE REFERENCE: PU3480US	
; CURRENT APPLICATION NUMBER: US/10/793,626	
; CURRENT FILING DATE: 2004-03-04	
; PRIOR APPLICATION NUMBER: 60/164,258	
; PRIOR FILING DATE: 1999-11-09	
; NUMBER OF SEQ ID NOS: 4472	
; SOFTWARE: PatentIn Ver. 2.1	
; SEQ ID NO 860	
; LENGTH: 618	
; TYPE: PRT	
; ORGANISM: Artificial Sequence	
; FEATURE:	
; OTHER INFORMATION: Description of Artificial Sequence: synthetic	
; OTHER INFORMATION: amino acid sequence	
US-10-793-626-860	
Query Match 4.5%; Score 90.5; DB 1; Length 618;	
Best Local Similarity 18.5%; Pred. No. 5.1;	
Matches 70; Conservative 50; Mismatches 130; Indels 129; Gaps 16;	
Qy 3 KSDFLSPKAIANRIKSGLOKLRWYQMCQKQCRDENGFKCHCMSESHORQILLASE--- 59	
Db 275 KGNVVDPNVLIDRY---GLDATRYVL-----MRELPGSDGVF 309	
Qy 60 NPOQPMDFSEFRNDPFLLELRREF--TKRVHNNIVNYEYISHREIHMMATQWETLTD 117	
Db 310 TPEAFVERTNYDLANDLNLNRTISMINKYFHGELPAYQGPKEHDEKMEAMALETVKS 369	
Qy 118 FT-----KWLGREGLCKVDETPKGWI----- 139	
Db 370 FNDNMESLQFSVALSTVWKFSIRTNK-YIDET-QPFWLAKDENQREMLGNVMAHLVENIR 427	
Qy 140 -----QVIDRDPETIRROLEKKKKQDLDDDEKTAFTIEEQVRRGLEKGEQETPVFT 192	
Db 428 FATILLQPPFLTHAPREIFKQLNINPNLDLHQLDSLQYQYMLSE-----AITVTEKPTPIFP 482	
Qy 193 ELSRENEEKVTFNLNKGNGSAGATTSSSLGPSALKLLGSAASGKR----- 241	
Db 483 RLDTAEATAYI-----KESMQPPKSIK--QSDPEPGKEQIDIKDFDKVE 523	
Qy 242 -KSSOSSQAQPAKKKKSALDEIMELEEEKRTARTDA-WLQP-----GIWVKIIT-----K 290	
Db 524 IKAATIIDAENVKSEKLLKIKVELDNEQRQIVSGIAKFPREDIIGKVAVVVNLKPAK 583	

Search completed: November 23, 2005, 16:39:51
Job time : 6.98725 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:33:44 ; Search time 132.661 Seconds
(without alignments)
1295.011 Million cell updates/sec

Title: US-09-555-529-25

Perfect score: 391

Sequence: 1 MGKSDPLSKAIANRIKSG.....GPLKGRVRGVIQVEDISKLA 391

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 2443163 seqs, 439378781 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A Geneseq 21:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1980s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*
- 9: Geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	371	94.9	391	2	AAY23619 Murine ki
2	179	45.8	393	2	AAY23620 A human k
3	179	45.8	393	7	ADJ69659 Human hea
4	163	41.7	291	2	AAY23617 Mouse tru
5	120	30.7	293	2	AAY23618 A huamn t
6	42	10.7	390	8	ADP22452 Sea-squir
7	35	9.0	411	3	AG42576 Arabidops
8	35	9.0	423	3	AAB03064 Maize KIN
9	35	9.0	423	3	AAB03065 Maize KIN
10	35	9.0	423	8	ADT58422 Plant pol
11	35	9.0	424	3	AAB03063 Maize KIN
12	35	9.0	437	8	ADX96318 Plant ful
13	35	9.0	437	8	ADX78817 Plant ful
14	29	7.4	84	4	AG474690 Human col
15	24	6.1	382	3	AG42577 Arabidops
16	21	5.4	302	3	AG46712 Arabidops
17	21	5.4	340	3	AG46711 Arabidops
18	21	5.4	343	3	AG46710 Arabidops
19	18	4.6	34	5	ADK35903 Novel hum
20	18	4.6	38	5	ADK35871 Novel hum
21	15	3.8	15	2	RA66765 RecA-like
22	15	3.8	390	4	ABB71416 Drosophil
23	11	2.8	11	2	AA66766 Zinc-fing
24	9	2.3	535	8	ADT60584 Plant pol

25	9	2.3	1542	5	ABE78013	Abb78013 Amino aci
26	9	2.3	3141	7	ADJ70444	Adj70444 Human hea
27	9	2.3	3144	2	AA58777	Aar58777 Protein e
28	9	2.3	3144	2	AAW36887	Aaw36887 Previousl
29	9	2.3	3144	2	AAW09871	Aaw09871 Human hun
30	9	2.3	3144	2	AAW44742	Aaw44742 Human hun
31	9	2.3	3144	2	AAY33493	Aay33493 Human hun
32	9	2.3	3144	9	ADY98141	Ady98141 Human hun
33	9	2.3	3223	4	ABB11407	Abb11407 Human Hun
34	9	2.3	3223	4	ABB11470	Abb11470 Human Hun
35	8	2.0	8	5	ABG96595	Abg96595 Human leu
36	8	2.0	8	5	ABG97218	Abg97218 Human leu
37	8	2.0	8	7	ADL99055	Adl99055 Human leu
38	8	2.0	8	7	ADL98437	Adl98437 Human com
39	8	2.0	10	4	AAG97042	Aag97042 Human com
40	8	2.0	159	3	AAG19456	Aag19456 Arabidops
41	8	2.0	173	4	AAU58581	Aau58581 Propionib
42	8	2.0	173	6	ABM55100	Abm55100 Propionib
43	8	2.0	179	8	ADJ93930	Adj93930 Human mit
44	8	2.0	183	4	ABB66749	Abb66749 Drosophil
45	8	2.0	192	4	ABB58711	Abb58711 Drosophil

ALIGNMENTS

RESULT 1
AAY23619
ID AAY23619 standard; protein; 391 AA.
XX
AC AAY23619;
DT 07-SEP-1999 (first entry)
XX
DE Murine kin17 protein.
XX
KW Mouse; kin17 protein; cell proliferation; fertility;
KW hyperproliferative disease; protein interaction; curved DNA;
KW HIV replication; HIV integration; repair enzyme.
XX
OS Mus sp.
XX
PN FR2772046-A1.
XX
PD 11-JUN-1999.
XX
PF 09-DEC-1997; 97FR-00015536.
XX
PR 09-DEC-1997; 97FR-00015536.
XX
(COMS) COMMISSARIAT ENERGIE ATOMIQUE.
XX
Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
WPI; 1999-359999/31.
XX
PT New DNA coding for human kin17 protein - useful for controlling cell
PT proliferation or fertility.
XX
PS Claim 19; Page 36-37; 69pp; French.
XX
CC The present sequence represents a murine kin17 protein. The mammalian
CC kin17 protein is useful for preparing a medicament for controlling cell
CC proliferation or for controlling fertility. The medicaments can also be
CC used to treat hyperproliferative diseases. Fragments between amino acids
CC 55 and 235 (preferably between amino acids 129 and 228) of a mammalian
CC kin17 protein are useful for regulating the interaction between proteins
CC and curved DNA. The fragment can be used to block replication of HIV or
CC its integration into the human genome or to target repair enzymes to
CC curved DNA sites. Expression vectors for kin17 can be used for
XX controlling cell proliferation
XX Sequence 391 AA;

Query Match	94.9%	Score 371;	DB 2;	Length 391;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 371;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	21	LQKLRWYQCMQKQCRDENGPKCHCMSESHORQLLASENPQQFMDYFSEEPNDFLELL	80	
Db	21	LQKLRWYQCMQKQCRDENGPKCHCMSESHORQLLASENPQQFMDYFSEEPNDFLELL	80	
Qy	81	RRRFGTKRVHNNI VYNEVISHREHIHMNATOWETLTDFTKMLGREGLCKVDETPKGWYIQ	140	
Db	81	RRRFGTKRVHNNI VYNEVISHREHIHMNATOWETLTDFTKMLGREGLCKVDETPKGWYIQ	140	
Qy	141	YIDRDPETIRRQLELEKKKKQDLDDDEEKTAKFIEBQVRRGLEGKEQETPVFTELSRENEE	200	
Db	141	YIDRDPETIRRQLELEKKKKQDLDDDEEKTAKFIEBQVRRGLEGKEQETPVFTELSRENEE	200	
Qy	201	EKVTFNLKNGAGGSAGATTSSKSSLGPSALKLLGSAASGKSKESQSSAQPAKKYKKSALD	260	
Db	201	EKVTFNLKNGAGGSAGATTSSKSSLGPSALKLLGSAASGKSKESQSSAQPAKKYKKSALD	260	
Qy	261	EIMELESEKKKTARTDAWLQPGIVVKIITTKLGEKYHKKKGVVKEVIDRYTAVVWMTDSG	320	
Db	261	EIMELESEKKKTARTDAWLQPGIVVKIITTKLGEKYHKKKGVVKEVIDRYTAVVWMTDSG	320	
Qy	321	DLRLKLDQTHLETVPAPGKRVLVNLNGGYRGNEGTLSEINAKFSATIVIEGTGPLKGRRVE	380	
Db	321	DLRLKLDQTHLETVPAPGKRVLVNLNGGYRGNEGTLSEINAKFSATIVIEGTGPLKGRRVE	380	
Qy	381	GIQYEDISKLA 391		
Db	381	GIQYEDISKLA 391		
RESULT 2				
ID	AA23620 standard; protein; 393 AA.			
XX	AA23620;			
XX	07-SEP-1999 (first entry)			
DT	A human kin17 protein.			
DE	Human; kin17 protein; cell proliferation; fertility;			
KW	Hyperproliferative disease; protein interaction; curved DNA;			
KW	HIV replication; HIV integration; repair enzyme.			
XX	Homo sapiens.			
XX	FR2772046-A1.			
PN	11-JUN-1999.			
PD	09-DEC-1997; 97FR-00015536.			
XX	09-DEC-1997; 97FR-00015536.			
PR	(COMS) COMMISSARIAT ENERGIE ATOMIQUE.			
XX	Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;			
XX	WPI; 1999-359999/31.			
DR	New DNA coding for human kin17 protein - useful for controlling cell			
XX	proliferation or fertility.			
PT	Claim 19; Page 37-38; 69pp; French.			
PS	The present sequence represents a human kin17 protein. The mammalian			
XX	kin17 protein is useful for preparing a medicament for controlling cell			
CC	proliferation or for controlling fertility. The medicaments can also be			
CC	used to treat hyperproliferative diseases. Fragments between amino acids			

CC altered mitochondrial function. Specifically, it refers to a method for
 CC identifying proteins of the human heart mitochondrial proteome that are
 CC useful for drug screening assays, as well as therapeutic targets. The
 CC present invention describes a method for identifying such proteins that
 CC can be used in the treatment of various diseases associated with altered
 CC mitochondrial function including diabetes mellitus, Huntington's disease,
 CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
 CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy
 CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
 CC compositions have neuroprotective, neurotropic, antidiabetic,
 CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and
 CC cyostatic activities. This polypeptide sequence is a human heart
 CC mitochondrial protein of the invention.

XX Sequence 393 AA;

SQ Query Match 45.0%; Score 179; DB 7; Length 393;
 Best Local Similarity 100.0%; Pred. No. 4.5e-164;
 Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 PKAIANRIKSGLOKLWYCOMQCKQCRDENGFKCHMSHQRQLLASENPQQFMDYF 68
 |||||
 DB 9 PKAIANRIKSGLOKLWYCOMQCKQCRDENGFKCHMSHQRQLLASENPQQFMDYF 68

QY 69 SEEFNDFLRLRRFCTKRVHNVVYNEVISHREIHMMATOWETLTDFTKLGRGLC 128
 |||||
 DB 69 SEEFNDFLRLRRFCTKRVHNVVYNEVISHREIHMMATOWETLTDFTKLGRGLC 128

QY 129 KVDETPKGWYIQYIDRDPETIRQLLEKKKKQDLDDDEKTAKFIEQVRRGLEKEQE 187
 |||||
 DB 129 KVDETPKGWYIQYIDRDPETIRQLLEKKKKQDLDDDEKTAKFIEQVRRGLEKEQE 187

RESULT 4

AAAY23617
 ID AAY23617 standard; protein; 291 AA.

XX AAY23617;

XX 07-SEP-1999 (first entry)

XX Mouse truncated kinl7 protein.

XX kinl7 protein; cell proliferation; fertility; hyperproliferative disease;
 KW protein interaction; curved DNA; HIV replication; HIV integration;
 KW repair enzyme.

XX Mus sp.

XX FR2772046-A1.

XX 11-JUN-1999.

XX 09-DEC-1997; 97FR-00015536.

XX 09-DEC-1997; 97FR-00015536.

XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;

XX WPI; 1999-359999/31.

XX New DNA coding for human kinl7 protein - useful for controlling cell
 PT proliferation or fertility.

XX Claim 14; Page 33-34; 69pp; French.

XX The present sequence represents a mouse kinl7 protein with amino acids
 CC 129-228 deleted. The mammalian kinl7 protein is useful for preparing a
 CC medicament for controlling cell proliferation or for controlling
 CC fertility. The medicaments can also be used to treat hyperproliferative
 CC diseases. Fragments between amino acids 55 and 235 (preferably between

CC amino acids 129 and 228) of a mammalian kinl7 protein are useful for
 CC regulating the interaction between proteins and curved DNA. The fragment
 CC can be used to block replication of HIV or its integration into the human
 CC genome or to target repair enzymes to curved DNA sites. Expression
 CC vectors for kinl7 can be used for controlling cell proliferation
 XX Sequence 291 AA;

SQ Query Match 41.7%; Score 163; DB 2; Length 291;
 Best Local Similarity 100.0%; Pred. No. 1e-148;
 Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 ALKLLGSAASGRKSSQSSAQPAPKKKSGALDEIMELEBEKKRTARTDAWLQPGIVVKII 288
 |||||
 DB 129 ALKLLGSAASGRKSSQSSAQPAPKKKSGALDEIMELEBEKKRTARTDAWLQPGIVVKII 188

QY 289 TKKLGEKHYKKGVKVEIDRYTAVVYKMTDSGRLKLDQTHLETVPAPGRVVLNGGY 348
 |||||
 DB 189 TKKLGEKHYKKGVKVEIDRYTAVVYKMTDSGRLKLDQTHLETVPAPGRVVLNGGY 248

QY 349 RGNEGTLESINEKAFSATIVETGPKGRRVGGIOVEDISKLA 391
 |||||

DB 249 RGNEGTLESINEKAFSATIVETGPKGRRVGGIOVEDISKLA 291

RESULT 5

AAAY23618
 ID AAY23618 standard; protein; 293 AA.

XX AAY23618;

XX 07-SEP-1999 (first entry)

XX A huamn truncated kinl7 protein.

XX kinl7 protein; cell proliferation; fertility; hyperproliferative disease;
 KW protein interaction; curved DNA; HIV replication; HIV integration;
 KW repair enzyme.

XX Homo sapiens.

XX FR2772046-A1.

XX 11-JUN-1999.

XX 09-DEC-1997; 97FR-00015536.

XX 09-DEC-1997; 97FR-00015536.

XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;

XX WPI; 1999-359999/31.

XX New DNA coding for human kinl7 protein - useful for controlling cell
 PT proliferation or fertility.

XX Claim 15; Page 34-35; 69pp; French.

XX The present sequence represents a truncated human kinl7 protein with
 CC amino acids 129-228 deleted. The mammalian kinl7 protein is useful for
 CC preparing a medicament for controlling cell proliferation or for
 CC controlling fertility. The medicaments can also be used to treat
 CC hyperproliferative diseases. Fragments between amino acids 55 and 235
 CC (preferably between amino acids 129 and 228) of a mammalian kinl7 protein
 CC are useful for regulating the interaction between proteins and curved
 CC DNA. The fragment can be used to block replication of HIV or its
 CC integration into the human genome or to target repair enzymes to curved
 CC DNA sites. Expression vectors for kinl7 can be used for controlling cell
 CC proliferation

XX Sequence 293 AA;

Query Match 30.7%; Score 120; DB 2; Length 293;
Best Local Similarity 100.0%; Pred. No. 4.3e-107;
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 PRAIANRIKSGLOKLRYWCQCKQCRDENGFKCHCMSESHQROLILLASENPQQFMDYF 68
Dy 9 PRAIANRIKSGLOKLRYWCQCKQCRDENGFKCHCMSESHQROLILLASENPQQFMDYF 68
Qy 69 SEEFNRDFLELLRRRFGTKRVHNNIVYNEYISHREHIHMNATQWETLDTFTKWLREGJLC 128
Dy 69 SEEFNRDFLELLRRRFGTKRVHNNIVYNEYISHREHIHMNATQWETLDTFTKWLREGJLC 128

RESULT 6
ADP22452
ID ADP22452 standard; protein; 390 AA.
AC ADP22452;
DT 12-AUG-2004 (first entry)
XX Sea-squirt (Ciona intestinalis) zinc finger protein #13.
DE sea-squirt; zinc finger protein; gene detection; drug development;
KW zinc finger protein-associated disease.
XX Ciona intestinalis.
OS JP2004057126-A.
XX 26-FEB-2004.
XX 31-JUL-2002; 2002JP-00222484.
XX 31-JUL-2002; 2002JP-00222484.
XX (KAGA-) KAGAKU GIUTSU SHINKO JIGYODAN.
XX WPI; 2004-208711/20.
DR N-PSDB; ADP22451.
XX Novel gene encoding zinc finger protein, useful as probe in gene
PT detecting instruments and in development of drug for treating zinc finger
PT protein associated diseases.
XX Claim 1; SEQ ID NO 26; 972pp; Japanese.
XX The invention comprises the amino acid and coding sequences of sea-squirt
CC (Ciona intestinalis) zinc finger proteins. The DNA and protein sequences
CC of the invention are useful in a gene detecting instrument. The DNA and
CC protein sequences of the invention are useful in the development of drugs
CC for the treatment of zinc finger protein-associated diseases. The present
CC amino acid sequence represents a sea-squirt zinc finger protein of the
CC invention.
XX Sequence 390 AA;
SQ

Query Match 10.7%; Score 42; DB 8; Length 390;
Best Local Similarity 100.0%; Pred. No. 1.8e-31;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 NRIKSGLOKLRYWCQCKQCRDENGFKCHCMSESHQROL 55
Dy 16 NRIKSGLOKLRYWCQCKQCRDENGFKCHCMSESHQROL 57

RESULT 7
AAG42576
ID AAG42576 standard; protein; 411 AA.
XX
AC AAG42576;
XX

DT 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 53114.
DE Arabidopsis thaliana protein fragment SEQ ID NO: 53114.
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
OS EP1033405-A2.
XX 06-SEP-2000.
PD
XX 25-FEB-2000; 2000EP-00301439.
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132484P.
PR 06-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.

PR	18-JUN-1999;	99US-0139763P.	PR	27-AUG-1999;	99US-0151066P.
PR	21-JUN-1999;	99US-0139817P.	PR	27-AUG-1999;	99US-0151080P.
PR	22-JUN-1999;	99US-0139899P.	PR	30-AUG-1999;	99US-0151303P.
PR	23-JUN-1999;	99US-0140353P.	PR	31-AUG-1999;	99US-0151438P.
PR	24-JUN-1999;	99US-0140354P.	PR	01-SEP-1999;	99US-0151930P.
PR	24-JUN-1999;	99US-0140695P.	PR	07-SEP-1999;	99US-0152363P.
PR	28-JUN-1999;	99US-0140823P.	PR	10-SEP-1999;	99US-0153070P.
PR	28-JUN-1999;	99US-0140891P.	PR	13-SEP-1999;	99US-0153758P.
PR	29-JUN-1999;	99US-0140921P.	PR	15-SEP-1999;	99US-0154018P.
PR	30-JUN-1999;	99US-0141287P.	PR	16-SEP-1999;	99US-0154039P.
PR	01-JUL-1999;	99US-0141842P.	PR	20-SEP-1999;	99US-0154779P.
PR	01-JUL-1999;	99US-0142154P.	PR	22-SEP-1999;	99US-0155139P.
PR	02-JUL-1999;	99US-0142055P.	PR	23-SEP-1999;	99US-0155486P.
PR	06-JUL-1999;	99US-0142390P.	PR	24-SEP-1999;	99US-0156559P.
PR	08-JUL-1999;	99US-0142803P.	PR	28-SEP-1999;	99US-0156458P.
PR	09-JUL-1999;	99US-0142820P.	PR	29-SEP-1999;	99US-0156596P.
PR	12-JUL-1999;	99US-0142977P.	PR	04-OCT-1999;	99US-0157117P.
PR	13-JUL-1999;	99US-0143542P.	PR	05-OCT-1999;	99US-0157753P.
PR	14-JUL-1999;	99US-0143624P.	PR	06-OCT-1999;	99US-0157865P.
PR	15-JUL-1999;	99US-0144005P.	PR	07-OCT-1999;	99US-0158029P.
PR	16-JUL-1999;	99US-0144085P.	PR	08-OCT-1999;	99US-0158232P.
PR	16-JUL-1999;	99US-0144086P.	PR	12-OCT-1999;	99US-0158369P.
PR	19-JUL-1999;	99US-0144325P.	PR	13-OCT-1999;	99US-0159293P.
PR	19-JUL-1999;	99US-0144331P.	PR	13-OCT-1999;	99US-0159294P.
PR	19-JUL-1999;	99US-0144332P.	PR	13-OCT-1999;	99US-0159295P.
PR	19-JUL-1999;	99US-0144333P.	PR	14-OCT-1999;	99US-0159329P.
PR	19-JUL-1999;	99US-0144334P.	PR	14-OCT-1999;	99US-0159330P.
PR	19-JUL-1999;	99US-0144335P.	PR	14-OCT-1999;	99US-0159331P.
PR	20-JUL-1999;	99US-0144352P.	PR	14-OCT-1999;	99US-0159637P.
PR	20-JUL-1999;	99US-0144632P.	PR	14-OCT-1999;	99US-0159638P.
PR	20-JUL-1999;	99US-0144884P.	PR	18-OCT-1999;	99US-0159584P.
PR	21-JUL-1999;	99US-0144814P.	PR	21-OCT-1999;	99US-0160741P.
PR	21-JUL-1999;	99US-0144866P.	PR	21-OCT-1999;	99US-0160767P.
PR	21-JUL-1999;	99US-01445088P.	PR	21-OCT-1999;	99US-0160768P.
PR	22-JUL-1999;	99US-01445085P.	PR	21-OCT-1999;	99US-0160770P.
PR	22-JUL-1999;	99US-0145087P.	PR	21-OCT-1999;	99US-0160814P.
PR	22-JUL-1999;	99US-0145089P.	PR	21-OCT-1999;	99US-0160815P.
PR	22-JUL-1999;	99US-0145192P.	PR	22-OCT-1999;	99US-0160980P.
PR	22-JUL-1999;	99US-0145193P.	PR	22-OCT-1999;	99US-0160981P.
PR	22-JUL-1999;	99US-0145276P.	PR	22-OCT-1999;	99US-0160989P.
PR	23-JUL-1999;	99US-0145224P.	PR	25-OCT-1999;	99US-0161404P.
PR	23-JUL-1999;	99US-0145218P.	PR	25-OCT-1999;	99US-0161405P.
PR	26-JUL-1999;	99US-0145913P.	PR	25-OCT-1999;	99US-0161406P.
PR	27-JUL-1999;	99US-0145918P.	PR	26-OCT-1999;	99US-0161359P.
PR	27-JUL-1999;	99US-0145919P.	PR	26-OCT-1999;	99US-0161360P.
PR	28-JUL-1999;	99US-0145951P.	PR	28-OCT-1999;	99US-0161361P.
PR	02-AUG-1999;	99US-0146386P.	PR	28-OCT-1999;	99US-0161920P.
PR	02-AUG-1999;	99US-0146388P.	PR	28-OCT-1999;	99US-0161922P.
PR	02-AUG-1999;	99US-0146389P.	PR	28-OCT-1999;	99US-0161993P.
PR	03-AUG-1999;	99US-0147038P.	PR	29-OCT-1999;	99US-0162142P.
PR	04-AUG-1999;	99US-0147204P.			
PR	04-AUG-1999;	99US-0147302P.			
PR	05-AUG-1999;	99US-0147192P.			
PR	05-AUG-1999;	99US-0147260P.			
PR	06-AUG-1999;	99US-0147303P.			
PR	06-AUG-1999;	99US-0147416P.			
PR	09-AUG-1999;	99US-0147493P.			
PR	09-AUG-1999;	99US-0147935P.			
PR	10-AUG-1999;	99US-0148171P.			
PR	11-AUG-1999;	99US-0148319P.			
PR	12-AUG-1999;	99US-0148341P.			
PR	13-AUG-1999;	99US-0148565P.			
PR	13-AUG-1999;	99US-0148684P.			
PR	16-AUG-1999;	99US-0149368P.			
PR	17-AUG-1999;	99US-0149175P.			
PR	18-AUG-1999;	99US-0149426P.			
PR	20-AUG-1999;	99US-0149722P.			
PR	20-AUG-1999;	99US-0149723P.			
PR	20-AUG-1999;	99US-0149929P.			
PR	23-AUG-1999;	99US-0149902P.			
PR	23-AUG-1999;	99US-0149930P.			
PR	25-AUG-1999;	99US-0150566P.			
PR	26-AUG-1999;	99US-0150884P.			
PR	27-AUG-1999;	99US-0151065P.			
Query Match					9.0%; Score 35; DB 3; Length 411;
Best Local Similarity					100.0%; Pred. No. 1.le-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
Qy	19	KGLQKLWYCMQCKQCRDENGFKCHCMSES	Qy	19	KGLQKLWYCMQCKQCRDENGFKCHCMSES
Db	19	KGLQKLWYCMQCKQCRDENGFKCHCMSES	Db	19	KGLQKLWYCMQCKQCRDENGFKCHCMSES
RESULT 8					
AAB03064					
ID	AAB03064 standard; protein; 423 AA.				
XX					
AC	AAB03064;				
XX					
DT	27-SEP-2000 (first entry)				
XX	Maize KIN17 orthologue, ZmKINH-2.				
DE					
XX	ZmKINH-2; KIN17 orthologue; maize; zinc finger protein; RecA homologue;				
KW	nuclear localisation; nonhomologous recombination;				
KW	illegitimate recombination; double stranded DNA binding; curved DNA;				
KW					

```

KW homologous gene targeting; transgenic plant.
XX
OS
XX
XX Zea mays.
XX WO200024900-A1.
XX PN
XX
XX
XX PD
XX 04-MAY-2000.
XX
XX PF 06-OCT-1999; 99WO-US023280.
XX
XX PR 27-OCT-1998; 98US-0105802P.
XX
XX PA (PION-) PIONEER HI-BRED INT INC.
XX
XX PI Mahajan PB;
XX
XX DR WPI; 2000-350741/30.
XX DR N-PSDB; AAA52591.
XX
XX PT Nucleic acids encoding maize KIN17 orthologue proteins useful for
XX preventing illegitimate recombination in cells.
XX
XX PS Claim 11; Page 64-66; 84pp; English.
XX
XX CC This sequence represents the maize KIN17 orthologue ZmKINH-2. The
XX invention relates to maize KIN17 orthologues (AAB03063-B03065), nucleic
XX acids encoding them (AAA52589-A52591), and expression vectors, transgenic
XX plants and plant seeds comprising nucleotides encoding maize KIN17
XX orthologues. KIN17 has, until now, been found only in animal (avian,
XX rodent and human) cells, this invention being the first report describing
XX the presence of KIN17 in plants. Murine KIN17 was found to have
XX significant homology to Escherichia coli RecA protein, and contains a
XX zinc finger motif and a nuclear localisation signal. KIN17 binds double-
XX stranded DNA, preferentially binding to curved DNA, and forms
XX intranuclear foci on overexpression in mammalian cells. It is also
XX induced on exposure to gamma or ultraviolet radiation. These findings
XX indicate that KIN17 plays a role in non-homologous (illegitimate)
XX recombination, which occurs at higher rates among higher eukaryotes,
XX particularly plants. Illegitimate recombination in plants is a major
XX impediment to the generation of transgenic crops such as maize. Maize
XX KIN17 orthologue nucleic acid sequences may be used to generate
XX transgenic plants. The transgenic plants generated can be monocots or
XX dicots and are particularly maize, soybean, sunflower, sorghum, canola,
XX wheat, alfalfa, cotton, rice, barley and millet. In particular, antisense
XX sequences may be used to reduce KIN17 levels in embryogenic callus or
XX embryo cells, thereby reducing the amount of non-homologous recombination
XX and enhancing homologous gene targeting
XX
XX SQ Sequence 423 AA;

Query Match 9.0%; Score 35; DB 3; Length 423;
Best Local Similarity 100.0%; Pred. No. 1.1e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRYWCQCKQCRDENGFKCHCMSESHQRQ 53
|||
DB 19 KGLQKLRYWCQCKQCRDENGFKCHCMSESHQRQ 53
|||

RESULT 9
AAB03065
ID AAB03065 standard; protein; 423 AA.
XX
XX AAB03065;
XX
XX 27-SEP-2000 (first entry)
XX
XX DE Maize KIN17 orthologue, ZmKINH-3.
XX
XX ZmKINH-3; KIN17 orthologue; maize; zinc finger protein; RecA homologue;
KW nuclear localisation; nonhomologous recombination;
KW illegitimate recombination; double stranded DNA binding; curved DNA;
KW homologous gene targeting; transgenic plant.
KW

```

```

XX
OS
XX
XX Zea mays.
XX WO200024900-A1.
XX PN
XX
XX
XX PD
XX 04-MAY-2000.
XX
XX PF 06-OCT-1999; 99WO-US023280.
XX
XX PR 27-OCT-1998; 98US-0105802P.
XX
XX PA (PION-) PIONEER HI-BRED INT INC.
XX
XX PI Mahajan PB;
XX
XX DR WPI; 2000-350741/30.
XX DR N-PSDB; AAA52591.
XX
XX PT Nucleic acids encoding maize KIN17 orthologue proteins useful for
XX preventing illegitimate recombination in cells.
XX
XX PS Claim 11; Page 64-66; 84pp; English.
XX
XX CC This sequence represents the maize KIN17 orthologue ZmKINH-3. The
XX invention relates to maize KIN17 orthologues (AAB03063-B03065), nucleic
XX acids encoding them (AAA52589-A52591), and expression vectors, transgenic
XX plants and plant seeds comprising nucleotides encoding maize KIN17
XX orthologues. KIN17 has, until now, been found only in animal (avian,
XX rodent and human) cells, this invention being the first report describing
XX the presence of KIN17 in plants. Murine KIN17 was found to have
XX significant homology to Escherichia coli RecA protein, and contains a
XX zinc finger motif and a nuclear localisation signal. KIN17 binds double-
XX stranded DNA, preferentially binding to curved DNA, and forms
XX intranuclear foci on overexpression in mammalian cells. It is also
XX induced on exposure to gamma or ultraviolet radiation. These findings
XX indicate that KIN17 plays a role in non-homologous (illegitimate)
XX recombination, which occurs at higher rates among higher eukaryotes,
XX particularly plants. Illegitimate recombination in plants is a major
XX impediment to the generation of transgenic crops such as maize. Maize
XX KIN17 orthologue nucleic acid sequences may be used to generate
XX transgenic plants. The transgenic plants generated can be monocots or
XX dicots and are particularly maize, soybean, sunflower, sorghum, canola,
XX wheat, alfalfa, cotton, rice, barley and millet. In particular, antisense
XX sequences may be used to reduce KIN17 levels in embryogenic callus or
XX embryo cells, thereby reducing the amount of non-homologous recombination
XX and enhancing homologous gene targeting
XX
XX SQ Sequence 423 AA;

Query Match 9.0%; Score 35; DB 3; Length 423;
Best Local Similarity 100.0%; Pred. No. 1.1e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRYWCQCKQCRDENGFKCHCMSESHQRQ 53
|||
DB 19 KGLQKLRYWCQCKQCRDENGFKCHCMSESHQRQ 53
|||

RESULT 10
ADT58422
ID ADT58422 standard; protein; 423 AA.
XX
XX ADT58422;
XX
XX 13-JAN-2005 (first entry)
XX
XX DE Plant polypeptide, SEQ ID 8499.
XX
XX Plant; transgenic; cold tolerance; growth rate; drought tolerance;
KW disease resistance; galactomannan production; plant growth regulator;
KW heat tolerance; herbicide tolerance; lignin production;
KW extreme osmotic condition tolerance; pathogens resistance;
KW pest resistance; yield improvement; seed oil yield; seed protein yield.
KW

```


ADX96318
ID ADX96318 standard; protein; 437 AA.
XX
AC ADX96318;
XX
DT 21-APR-2005 (first entry)
XX
DE Plant full length insert polypeptide seqid 58982.
XX
KW plant protectant; plant growth regulant; gene therapy; plant;
KW recombinant DNA construct; physical array; plant breeding marker;
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW extreme osmotic condition; pathogen tolerance; pest tolerance;
KW growth rate; cell cycle pathway; disease resistance;
KW galactomannan production; lignin production; plant growth regulator;
KW yield; plant growth; plant development; seed oil; protein yield;
KW protein content.
XX
OS Unidentified.
XX
PN US2004034888-A1.
XX
PD 19-FEB-2004.
XX
PF 28-APR-2003; 2003US-00425114.
XX
PR 06-MAY-1999; 99US-00304517.
PR 05-NOV-2001; 2001US-00985678.
XX
PA (LIUJ/) LIU J.
PA (ZHOU/) ZHOU Y.
PA (KOVA/) KOVALIC D K.
PA (SCRE/) SCREEN S E.
PA (TABA/) TABASKA J E.
PA (CAOY/) CAO Y.
PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
XX WPI; 2004-180133/17.
XX
New recombinant DNA construct, useful for improving plant tolerance to
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT pests, for conferring increased resistance to plant disease, or for
PT improving yield.
XX
Claim 1; SEQ ID NO 58982; 15pp; English.
XX
The invention describes a recombinant DNA construct comprising a
CC polynucleotide consisting of a sequence encoding an amino acid sequence
CC available in electronic form from the US patent office at
CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC increased resistance to plant disease, for producing galactomannan,
CC lignin or plant growth regulators, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This is the amino acid sequence of a plant full length insert
CC polypeptide that can be used in the recombinant DNA construct of the
XX invention.
XX
SQ Sequence 437 AA;
Query Match 9.0%; Score 35; DB 8; Length 437;
Best Local Similarity 100.0%; Pred. No. 1.2e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

||||| 32 KGLQLRWYQMCQKQCRDENGFKCHCMSESHQRQ 66
DB
RESULT 13
ADX78817
ID ADX78817 standard; protein; 437 AA.
XX
AC ADX78817;
XX
DT 21-APR-2005 (first entry)
XX
DE Plant full length insert polypeptide seqid 48183.
XX
KW plant protectant; plant growth regulant; gene therapy; plant;
KW recombinant DNA construct; physical array; plant breeding marker;
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW extreme osmotic condition; pathogen tolerance; pest tolerance;
KW growth rate; cell cycle pathway; disease resistance;
KW galactomannan production; lignin production; plant growth regulator;
KW yield; plant growth; plant development; seed oil; protein yield;
KW protein content.
XX
OS Unidentified.
XX
PN US2004034888-A1.
XX
PD 19-FEB-2004.
XX
PF 28-APR-2003; 2003US-00425114.
XX
PR 06-MAY-1999; 99US-00304517.
PR 05-NOV-2001; 2001US-00985678.
XX
PA (LIUJ/) LIU J.
PA (ZHOU/) ZHOU Y.
PA (KOVA/) KOVALIC D K.
PA (SCRE/) SCREEN S E.
PA (TABA/) TABASKA J E.
PA (CAOY/) CAO Y.
PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
XX WPI; 2004-180133/17.
XX
New recombinant DNA construct, useful for improving plant tolerance to
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT pests, for conferring increased resistance to plant disease, or for
PT improving yield.
XX
Claim 1; SEQ ID NO 48183; 15pp; English.
XX
The invention describes a recombinant DNA construct comprising a
CC polynucleotide consisting of a sequence encoding an amino acid sequence
CC available in electronic form from the US patent office at
CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC increased resistance to plant disease, for producing galactomannan,
CC lignin or plant growth regulators, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This is the amino acid sequence of a plant full length insert
CC polypeptide that can be used in the recombinant DNA construct of the
XX invention.
XX
SQ Sequence 437 AA;

Query Match 9.0%; Score 35; DB 8; Length 437;
Best Local Similarity 100.0%; Pred. No. 1.2e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLWYQMCQKQCRDENGPKCHCMSESHQK 53
DB 32 KGLQKLWYQMCQKQCRDENGPKCHCMSESHQK 66

RESULT 14
AAG74690
ID AAG74690 standard; protein; 84 AA.

XX AAG74690;
AC AAG74690;
XX 03-SEP-2001 (first entry)
DT DT
XX Human colon cancer antigen protein SEQ ID NO:5454.
DE Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma.
KW Homo sapiens.
XX WO200122920-A2.
XX 05-APR-2001.
XX 28-SEP-2000; 2000WO-US026524.
XX 29-SEP-1999; 99US-0157137P.
XX 03-NOV-1999; 99US-0163280P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX WPI; 2001-235357/24.
XX N-PSDB; AAH34095.

CC Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers.
PS Claim 11; Page 7064-7065; 9803pp; English.
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene therapy
CC and vaccine production. N and P may be used in the prevention, diagnosis
CC and treatment of diseases associated with inappropriate P expression. For
CC example, N and P may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of P by expressing inactive proteins or to
CC supplement the patients own production of P. Additionally, N may be used
CC to produce the colon cancer-associated Ps, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the proteins. N and P
CC can be used in the prevention, diagnosis and treatment of colorectal
CC carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent
CC sequences used in the exemplification of the present invention. N.B.
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC time of publication, meaning no sequences are present for SEQ ID NO:1027
CC to 1052, 7921 and 7922

Sequence 84 AA;

Query Match 7.4%; Score 29; DB 4; Length 84;
Best Local Similarity 100.0%; Pred. No. 1.6e-19;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 363 FSATIVETGPKRRVGGIYEDISKLA 391
DB 56 FSATIVETGPKRRVGGIYEDISKLA 84

RESULT 15
AAG42577
ID AAG42577 standard; protein; 382 AA.

XX AAG42577;
AC AAG42577;
XX 18-OCT-2000 (first entry)
DT DT
XX Arabidopsis thaliana protein fragment SEQ ID NO: 53115.
DE Arabidopsis thaliana protein fragment SEQ ID NO: 53115.
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125788P.
XX 25-MAR-1999; 99US-0126264P.
XX 29-MAR-1999; 99US-0126785P.
XX 01-APR-1999; 99US-0127462P.
XX 06-APR-1999; 99US-0128234P.
XX 08-APR-1999; 99US-0128714P.
XX 16-APR-1999; 99US-0129845P.
XX 19-APR-1999; 99US-0130077P.
XX 21-APR-1999; 99US-0130449P.
XX 23-APR-1999; 99US-0130510P.
XX 28-APR-1999; 99US-0130891P.
XX 30-APR-1999; 99US-0131449P.
XX 30-APR-1999; 99US-0132048P.
XX 04-MAY-1999; 99US-0132407P.
XX 05-MAY-1999; 99US-0132484P.
XX 06-MAY-1999; 99US-0132485P.
XX 08-MAY-1999; 99US-0132486P.
XX 08-MAY-1999; 99US-0132487P.
XX 07-MAY-1999; 99US-0132863P.
XX 11-MAY-1999; 99US-0134256P.
XX 14-MAY-1999; 99US-0134218P.
XX 14-MAY-1999; 99US-0134219P.
XX 14-MAY-1999; 99US-0134221P.
XX 14-MAY-1999; 99US-0134370P.
XX 18-MAY-1999; 99US-0134941P.
XX 19-MAY-1999; 99US-0135124P.
XX 20-MAY-1999; 99US-0135353P.
XX 21-MAY-1999; 99US-0135629P.
XX 23-MAY-1999; 99US-0136021P.
XX 27-MAY-1999; 99US-0136392P.
XX 28-MAY-1999; 99US-0136782P.
XX 01-JUN-1999; 99US-0137222P.
XX 03-JUN-1999; 99US-0137528P.
XX 04-JUN-1999; 99US-0137502P.
XX 07-JUN-1999; 99US-0137724P.
XX 08-JUN-1999; 99US-0138094P.
XX 10-JUN-1999; 99US-0138540P.
XX 10-JUN-1999; 99US-0138847P.
XX 14-JUN-1999; 99US-0139119P.
XX 16-JUN-1999; 99US-0139452P.
XX 16-JUN-1999; 99US-0139453P.
XX 17-JUN-1999; 99US-0139492P.
XX 18-JUN-1999; 99US-0139454P.
XX 18-JUN-1999; 99US-0139455P.
XX 18-JUN-1999; 99US-0139456P.

PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 28-JUL-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 05-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.

PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154019P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 6.1%; Score 24; DB 3; Length 382;

Best Local Similarity 100.0%; Pred. No. 4.5e-14; Mismatches 0; Indels 0; Gaps 0;

Qy 30 MCQKQCRDENGFKCHCMSESHQRQ 53

Db 1 MCQKQCRDENGFKCHCMSESHQRQ 24

Search completed: November 23, 2005, 16:44:22
Job time : 134.661 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

QM protein - protein search, using sw model

Run on: November 23, 2005, 16:39:45 ; Search time 24.9362 Seconds
(without alignments)
1508.679 Million cell updates/sec

Title: US-09-555-529-25
Perfect score: 391
Sequence: 1 MGKSDFLSPKAIANRIKSG.....GPLKGRVVGIIQYEDISKLA 391

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 0
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 80.*
1: Piri.*
2: Piri2.*
3: Piri3.*
4: Piri4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	367	93.9	391	2 S18666	KIN17 protein - mo
2	35	9.0	411	2 H96596	hypothetical prote
3	10	2.6	404	2 T27106	hypothetical prote
4	9	2.3	304	2 T40316	conserved kin7-lik
5	9	2.3	3144	2 A46068	Huntington disease
6	9	2.0	303	1 S75983	hypothetical prote
7	8	2.0	342	2 A48258	dopamine receptor
8	8	2.0	400	2 G00013	D3 dopamine recept
9	8	2.0	400	2 G01977	d3 dopamine recept
10	8	2.0	446	1 DVRTD3	dopamine receptor
11	8	2.0	446	2 I48322	dopamine receptor
12	8	2.0	461	2 F05711	hypothetical glyci
13	8	2.0	514	2 D56849	receptor
14	8	2.0	802	2 C84733	probable O-GlcNAc
15	8	2.0	808	2 G86185	hypothetical prote
16	8	2.0	937	1 S42366	endopeptidase La h
17	8	2.0	962	1 S73342	endopeptidase La h
18	8	2.0	1901	2 F70806	hypothetical glyci
19	7	1.8	57	2 A12525	transposase asr738
20	7	1.8	70	2 B10239	hypothetical prote
21	7	1.8	89	2 AF2709	hypothetical prote
22	7	1.8	89	2 G97491	hypothetical prote
23	7	1.8	100	2 D82245	methylated-DNA-pro
24	7	1.8	121	2 D75089	hypothetical prote
25	7	1.8	136	2 AB2542	transcription repr
26	7	1.8	138	1 Q0BE82	UL73 glycoprotein
27	7	1.8	144	2 C81093	hypothetical prote
28	7	1.8	150	2 A55209	H transfer determi
29	7	1.8	189	2 A80744	conserved hypothet

hypothetical prote
hypothetical prote
probable transaldo
probable transaldo
probable transaldo
transaldolase-like
phosphoglyceromuta
hypothetical prote
vancomycin resista
hypothetical prote
hypothetical prote
spectrin alpha cha
hypothetical prote
merozoite surface
merozoite surface
hypothetical prote

ALIGNMENTS

RESULT 1

S18666
KIN17 protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 19-Feb-1994 #sequence revision 10-Nov-1995 #text_change 09-Jul-2004
C:Accession: S18666; A43753; S14622
R:Angulo, J.F.; Rouer, E.; Mazin, A.; Mattei, M.G.; Tissier, A.; Horellou, P.; Benatrous, R.
Nucleic Acids Res. 19, 5117-5123, 1991
A:Title: Identification and expression of the cDNA of KIN17, a zinc-finger gene located c
A:Reference number: S18666; MUID:92020193; PMID:1923796
A:Accession: S18666
A:Molecule type: mRNA
A:Residues: 1-391 <ANG>
A:Cross-references: UNIPROT:Q8K339; UNIPARC:UPI000017993F; EMBL:X58472
R:Angulo, J.; Rouer, E.; Benatrous, R.; Devoret, R.
Biochimie 73, 251-256, 1991
A:Title: Identification of a mouse cDNA fragment whose expressed polypeptide reacts with
A:Reference number: A43753; MUID:91355299; PMID:1715759
A:Accession: A43753
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 74-273 <ANW>
A:Cross-references: UNIPARC:UPI0000179940
C:Superfamily: KIN17 protein

Query Match	93.9%	Score 367;	DB 2;	Length 391;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 367;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	25	RWYQCMQKCRDENGFKCHCMSESHORQLLLASENPQDFMDYFSEFRNDFLELLRRRF	84	
Db	25	RWYQCMQKCRDENGFKCHCMSESHORQLLLASENPQDFMDYFSEFRNDFLELLRRRF	84	
Qy	85	GTRKRVHNNIVYNEVISHREIHNNATQWETLTDTFTKWLREGLCCKVDETPKGMYIQYIDR	144	
Db	85	GTRKRVHNNIVYNEVISHREIHNNATQWETLTDTFTKWLREGLCCKVDETPKGMYIQYIDR	144	
Qy	145	DPETIRROLEKKKKQDLDEEKTAKFIEQVRRRGLEGKEQETPVFTELSENREEBEKVT	204	
Db	145	DPETIRROLEKKKKQDLDEEKTAKFIEQVRRRGLEGKEQETPVFTELSENREEBEKVT	204	
Qy	205	FNLNKGAGGAGATTSSSSSLGPSALKLGSASGKKESSQSAQPAKKKKSALDEIME	264	
Db	205	FNLNKGAGGAGATTSSSSSLGPSALKLGSASGKKESSQSAQPAKKKKSALDEIME	264	
Qy	265	LEEEKKRTATDQWLPQGIIVVKIITKLGKHYHKKGVKVEVIDRYTAVVKMTDSGDRLK	324	
Db	265	LEEEKKRTATDQWLPQGIIVVKIITKLGKHYHKKGVKVEVIDRYTAVVKMTDSGDRLK	324	
Qy	325	LDQTHLETVPAPGKRVLVNGGYRGNEGTLESINEKAFSATIVTETGPKLGRVEGIIQY	384	
Db	325	LDQTHLETVPAPGKRVLVNGGYRGNEGTLESINEKAFSATIVTETGPKLGRVEGIIQY	384	

QY 385 EDISKLA 391
Db 385 EDISKLA 391

RESULT 2
H96596
hypothetical protein TSA14.13 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: H96596
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizär, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: H96596
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-411 <STO>
A:Cross-references: UNIPROT:Q92VU5; UNIPARC:UPI00000A3CC8; GB:AE005173; NID:g4204268; PI
C:Genetics:
A:Gene: TSA14.13
A:Map position: 1
C:Superfamily: KIN17 protein

Query Match 9.0%; Score 35; DB 2; Length 411;
Best Local Similarity 100.0%; Pred. No. 1.8e-26;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLWWCQMKQKQCRDENGFKCHCMSESHQRQ 53
Db 19 KGLQKLWWCQMKQKQCRDENGFKCHCMSESHQRQ 53

RESULT 3
T27106
hypothetical protein Y52B11A.9 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T27106
R:Lennard, N.
submitted to the EMBL Data Library, October 1998
A:Reference number: Z20310
A:Accession: T27106
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-404 <WL>
A:Cross-references: UNIPROT:Q9XWF2; UNIPARC:UPI0000078B04; EMBL:AL032654; PIDN:CAA21720.
A:Experimental source: clone Y52B11A
C:Genetics:
A:Gene: CESP:Y52B11A.9
A:Map position: 1
A:Introns: 26/2; 102/3; 301/3; 367/3
C:Superfamily: KIN17 protein

Query Match 2.6%; Score 10; DB 2; Length 404;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 CQMCQKQCRD 37
Db 28 CQMCQKQCRD 37

RESULT 4

T40316
conserved kin7-like hypothetical protein SPBC365.09c - fission yeast (Schizosaccharomycet
C:Species: Schizosaccharomycetes pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T40316
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Lelaure, V.; Galibert, F.
submitted to the EMBL Data Library, June 1999
A:Reference number: Z21920
A:Accession: T40316
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-304 <WOO>
A:Cross-references: UNIPROT:Q9YX9; UNIPARC:UPI000006B457; EMBL:AL078627; PIDN:CAB44761.1
A:Experimental source: strain 972h; cosmid c365
C:Genetics:
A:Gene: SPDB:SPBC365.09c
A:Map position: 2

Query Match 2.3%; Score 9; DB 2; Length 304;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 RDENGFKCH 44
Db 36 RDENGFKCH 44

RESULT 5
A46068
Huntington disease-associated protein - human
C:Species: Homo sapiens (man)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Jul-2004
C:Accession: A46068; I54337
R:MacDonald, M.B.; Ambrose, C.M.; Duyao, M.P.; Myers, R.H.; Lin, C.; Srinidhi, L.; Barnes
J.P.; Bates, G.P.; Baxendale, S.; Hummerich, H.; Kirby, S.; North, M.; Youngman, S.; Mott
O'Donovan, M.C.; Riba-Ramirez, L.; Shah, M.; Stanton, V.P.; Strobel, S.A.; Draths, K.M.
Cell 72, 971-983, 1993
A:Authors: Wales, J.L.; Dervan, P.; Housman, D.E.; Altherr, M.; Shiang, R.; Thompson, L.;
d, K.; Collins, F.S.; Snell, R.; Holloway, T.; Gillespie, K.; Datson, N.; Shaw, D.; Harpe
A:Title: A novel gene containing a trinucleotide repeat that is expanded and unstable on
A:Reference number: A46068; MUID:93208892; PMID:8458085
A:Accession: A46068
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-3144 <MAC>
A:Cross-references: UNIPROT:P42858; UNIPARC:UPI0000146F05; GB:U12392
R:Lin, B.; Rommens, J.M.; Graham, R.K.; Kalchman, M.; MacDonald, H.; Nasir, J.; Delaney,
Hum. Mol. Genet. 2, 1541-1545, 1993
A:Title: Differential 3' polyadenylation of the Huntington disease gene results in two m
A:Reference number: I54337; MUID:94093536; PMID:7903579
A:Accession: I54337
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 2563-3144 <RES>
A:Cross-references: UNIPARC:UPI000016AABF; GB:L20431; NID:g398028; PIDN:AAA52702.1; PID:
C:Genetics:
A:Gene: GDB:HD
A:Cross-references: GDB:119307; OMIM:143100
A:Map position: 4p16.3-4p16.3

Query Match 2.3%; Score 9; DB 2; Length 3144;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 TTSKSSSLG 226
Db 1222 TTSKSSSLG 1230

RESULT 6
S75983
hypothetical protein - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.

A:Variety: PCC 6803
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
R:Accession: S75983
C:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpō, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S75983
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-303 <KAN>
A:Cross-references: UNIPROT:Q55483; UNIPARC:UPI000000C0EFA; EMBL:D64006; GB:AB001339; NID submitted to the EMBL Data Library, June 1996
C:Superfamily: hypothetical protein ybbp

Query Match 2.0%; Score 8; DB 1; Length 303;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 76 FLELLRR 83
| | | | |
Db 100 FLELLRR 107

RESULT 7
A48258
dopamine receptor variant D3nf (schizophrenia) - human
C:Species: Homo sapiens (man)
C:Date: 24-May-1996 #sequence_revision 24-May-1996 #text_change 09-Jul-2004
R:Schmauss, C.; Haroutunian, V.; Davis, K.L.; Davidson, M.
Proc. Natl. Acad. Sci. U.S.A. 90, 8942-8946, 1993
A:Title: Selective loss of dopamine D3-type receptor mRNA expression in parietal and motor cortex in schizophrenia
A:Reference number: A48258; MUID:94022291; PMID:8415635
A:Accession: A48258
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-342 <RES>
A:Cross-references: UNIPROT:P35462; UNIPARC:UPI000002A7A9; GB:L20469; NID:G306688; PIDN: submitted to the EMBL Data Library, February 1995
C:Superfamily: neurotransmitter receptor
C:Keywords: neurotransmitter receptor

Query Match 2.0%; Score 8; DB 2; Length 342;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 307 IDRYTAVV 314
| | | | |
Db 126 IDRYTAVV 133

RESULT 8
G00013
D3 dopamine receptor - green monkey
C:Species: Cercopithecus aethiops (green monkey, grivet)
C:Date: 13-Mar-1997 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
C:Accession: G00013
R:Ross, P.C.
submitted to the EMBL Data Library, February 1995
A:Reference number: G00049
A:Accession: G00013
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-400 <ROS>
A:Cross-references: UNIPROT:P52703; UNIPARC:UPI00000128DB0; EMBL:U211307; NID:G984965; PID submitted to the EMBL Data Library, February 1995
C:Superfamily: neurotransmitter receptor
C:Keywords: neurotransmitter receptor

Query Match 2.0%; Score 8; DB 2; Length 400;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 307 IDRYTAVV 314
| | | | |
Db 126 IDRYTAVV 133

RESULT 9

G01977
d3 dopamine receptor - human
C:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 09-Jul-2004
C:Accession: G01977
R:Fishburn, C.S.; Park, B.
submitted to the EMBL Data Library, July 1995
A:Reference number: G08971
A:Accession: G01977
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-400 <FIS>
A:Cross-references: UNIPROT:P35462; UNIPARC:UPI000002CDEFB; EMBL:U32499; NID:G927341; PIDN: submitted to the EMBL Data Library, July 1995
C:Superfamily: vertebrate rhodopsin

Query Match 2.0%; Score 8; DB 2; Length 400;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 307 IDRYTAVV 314
| | | | |
Db 126 IDRYTAVV 133

RESULT 10

DYRTD3
dopamine receptor D3 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 09-Jul-2004
C:Accession: S11565; S41849; I52280
R:Sokoloff, P.; Giros, B.; Martres, M.P.; Bouthenet, M.L.; Schwartz, J.C.
Nature 347, 146-151, 1990
A:Title: Molecular cloning and characterization of a novel dopamine receptor (D3) as a member of the G-protein-coupled receptor family
A:Reference number: S11565; MUID:90370111; PMID:1975644
A:Accession: S11565
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-446 <SOK>
A:Cross-references: UNIPROT:P19020; UNIPARC:UPI000002CB23; EMBL:X53944; NID:G56060; PIDN: submitted to the EMBL Data Library, February 1995
R:Giros, B.; Martres, M.P.; Pilon, C.; Sokoloff, P.; Schwartz, J.C.
Biochem. Biophys. Res. Commun. 176, 1584-1592, 1991
A:Title: Shorter variants of the D(3) dopamine receptor produced through various pattern mutations
A:Reference number: S41849; MUID:91248260; PMID:2039532
A:Accession: S41849
A:Molecule type: DNA
A:Residues: 1-138, 'E', 140-446 <GIR>
A:Cross-references: UNIPARC:UPI000002CB22
R:Pagliusi, S.; Chollet-Daemmerius, A.; Losberger, C.; Mills, A.; Kawashima, E.
Biochem. Biophys. Res. Commun. 194, 465-471, 1993
A:Title: Characterization of a novel exon within the D3 receptor gene giving rise to an alternative splicing variant
A:Reference number: I52280; MUID:93326159; PMID:8333859
A:Accession: I52280
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 71-106 <RES>
A:Cross-references: UNIPARC:UPI00000170C6A; GB:S63847; NID:G399705; PIDN:AAB27545.1; PID: submitted to the EMBL Data Library, February 1995
A:Experimental source: strain Wistar, brain
C:Genetics:
A:Introns: 90/3
C:Superfamily: vertebrate rhodopsin
C:Keywords: G protein-coupled receptor; glycoprotein; lipoprotein; neurotransmitter receptor

F:33-55/Domain: transmembrane #status predicted <TM1>
F:67-92/Domain: transmembrane #status predicted <TM2>
F:105-126/Domain: transmembrane #status predicted <TM3>
F:150-172/Domain: transmembrane #status predicted <TM4>
F:186-209/Domain: transmembrane #status predicted <TM5>

```

F;210-375/Domain: intracellular #status predicted <CYT>
F;376-399/Domain: transmembrane #status predicted <TM6>
F;413-434/Domain: transmembrane #status predicted <TM7>
F;12,19/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;103-181/Disulfide bonds: #status predicted
F;446/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match                2.0%; Score 8; DB 1; Length 446;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 IDRYTAVV 314
Db 126 IDRYTAVV 133
|||||

RESULT 11
I48322
dopamine receptor D3 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 09-Jul-2004
C;Accession: I48322; S25161
R;Fishburn, C.S.; Belleli, D.; David, C.; Carmon, S.; Fuchs, S.
J. Biol. Chem. 268, 5872-5878, 1993
A;Title: A novel short isoform of the D3 dopamine receptor generated by alternative splicing
A;Reference number: I48322; MUID:93194894; PMID:8449953
A;Accession: I48322
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-446 <RES>
A;Cross-references: UNIPROT:P30728; UNIPARC:UPI0000021FB7; EMBL:X67274; NID:G50651; PTDN
C;Superfamily: vertebrate rhodopsin
C;Keywords: neurotransmitter receptor

Query Match                2.0%; Score 8; DB 2; Length 446;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 IDRYTAVV 314
Db 126 IDRYTAVV 133
|||||

RESULT 12
F70571
hypothetical glycine-rich protein RV2615c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
C;Accession: F70571
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: F70571
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-461 <COL>
A;Cross-references: UNIPROT:O06199; UNIPARC:UPI0000165306; GB:Z95387; GB:AL123456; NID:9
A;Experimental source: strain H37RV
C;Genetics:
A;Gene: RV2615c
C;Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match                2.0%; Score 8; DB 2; Length 461;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 GAGGSAGA 217
Db 379 GAGGSAGA 386
|||||

```

RESULT 13

D56849

dopamine receptor-like protein D222 - Japanese pufferfish

C;Species: Fugu rubripes (Japanese pufferfish)

C;Date: 27-Oct-1995 #sequence_revision 27-Oct-1995 #text_change 09-Jul-2004

C;Accession: D56849

R;Macrae, A.D.; Brenner, S.

Genomics 25, 436-446, 1995

A;Title: Analysis of the dopamine receptor family in the compact genome of the puffer fish

A;Reference number: A56849; MUID:95309911; PMID:7789977

A;Accession: D56849

A;Status: preliminary; not compared with conceptual translation

A;Molecule type: DNA

A;Residues: 1-514 <MAC>

A;Cross-references: UNIPROT:O9PSA6; UNIPARC:UPI00000FBA2D

C;Superfamily: vertebrate rhodopsin

C;Keywords: neurotransmitter receptor

Query Match 2.0%; Score 8; DB 2; Length 514;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 IDRYTAVV 314

Db 131 IDRYTAVV 138

|||||

RESULT 14

C84733

probable O-GlcNAc transferase [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004

C;Accession: C84733

R;Lin, X.; Kaul, S.; Rounseley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; A.

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: C84733

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-802 <STO>

A;Cross-references: UNIPROT:Q8S8L9; UNIPARC:UPI000000C2F7; GB:AE002093; NID:G6598434; PII

C;Genetics:

A;Gene: At2g32450

A;Map position: 2

Query Match 2.0%; Score 8; DB 2; Length 802;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 ALKLLGSA 236

Db 341 ALKLLGSA 348

|||||

RESULT 15

G86185

hypothetical protein [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004

C;Accession: G86185

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Hultzar, L.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luross, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, J.

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: G86185
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-808 <STO>
A:Cross-references: UNIPROT:O23052; UNIPARC:UPI00000481C8; GB:AE005172; NID:g2388582; PI
C:Genetics:
A:Map position: 1

Query Match 2.0%; Score 8; DB 2; Length 808;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 229 ALKILGSA 236
Db 346 ALKILGSA 353
|||||

Search completed: November 23, 2005, 16:51:46
Job time : 26.9362 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:34:39 ; Search time 154.106 Seconds
(without alignments)
1790.078 Million cell updates/sec

Title: US-09-555-529-25

Perfect score: 391

Sequence: 1 MGKSDFLSPKAIANRIKSG.....GPLKGRRVGIQVEDISKLA 391

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 2166443 seqs, 705528306 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Uniprot 05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	391	100.0	391	2	Q8K339_MOUSE	Q8K339 mus musculus
2	195	49.9	195	2	Q9CV58_MOUSE	Q9CV58 mus musculus
3	179	45.8	393	2	Q60870_HUMAN	Q60870 homo sapien
4	61	15.6	389	2	Q6G197_XENTR	Q6G197 xenopus tro
5	50	12.8	387	2	Q68F56_XENLA	Q68F56 xenopus lae
6	48	12.3	427	2	Q4S5G2_TETNG	Q4S5G2 tetraodon n
7	43	11.0	383	2	Q5RH25_BRARE	Q5RH25 brachydanio
8	41	10.5	269	2	Q4TDV9_TETNG	Q4TDV9 tetraodon n
9	35	9.0	402	2	Q9AY87_ORYSA	Q9AY87 oryza sativ
10	35	9.0	411	2	Q9ZVU5_ARATH	Q9ZVU5 arabidopsis
11	35	9.0	430	2	Q75LU5_ORYSA	Q75LU5 oryza sativ
12	20	5.1	188	2	Q4TDW0_TETNG	Q4TDW0 tetraodon n
13	15	3.8	178	2	Q6X190_DROVA	Q6X190 drosophila
14	15	3.8	244	2	Q76926_DROME	Q76926 drosophila
15	15	3.8	390	2	Q8SXK2_DROME	Q8SXK2 drosophila
16	15	3.8	390	2	Q9VPH4_DROME	Q9VPH4 drosophila
17	14	3.6	263	2	Q5CFM3_CRYHO	Q5CFM3 cryptospori
18	14	3.6	265	2	Q5CYD1_CRYPV	Q5CYD1 cryptospori
19	13	3.3	387	2	Q7PGAL_ANOGA	Q7PGAL anopheles g
20	12	3.1	121	2	Q4XND8_PLACH	Q4XND8 plasmodium
21	12	3.1	239	2	Q6CXJ7_KLUFA	Q6CXJ7 kluyveromyc
22	12	3.1	378	2	Q4N9K7_THEPA	Q4N9K7 theileria p
23	12	3.1	412	2	Q4UG63_THEAN	Q4UG63 theileria a
24	12	3.1	441	2	Q4XDV6_PLACH	Q4XDV6 plasmodium
25	12	3.1	442	2	Q8IKG2_PLAF7	Q8IKG2 plasmodium
26	12	3.1	445	2	Q5SD16_DICDI	Q5SD16 dictyosteli
27	12	3.1	445	2	Q4Z7D7_PLABE	Q4Z7D7 plasmodium
28	12	3.1	457	2	Q7RD49_PLAYO	Q7RD49 plasmodium
29	11	2.8	238	2	Q51F62_ENTHI	Q51F62 entamoeba h
30	10	2.6	397	2	Q60L39_CAEBR	Q60L39 caenorhabdi
31	10	2.6	404	2	Q9XWF2_CABEL	Q9XWF2 caenorhabdi

32	9	2.3	304	2	Q9Y7X9_SCHPO	Q9Y7X9 echizoaecch
33	9	2.3	324	2	Q55M94_CRYNE	Q55M94 cryptococcu
34	9	2.3	324	2	Q5K8H7_CRYNE	Q5K8H7 cryptococcu
35	9	2.3	334	2	Q4WFG0_ASPFU	Q4WFG0 aspergillus
36	9	2.3	335	2	Q5AQSO_EMENI	Q5AQSO aspergillus
37	9	2.3	1147	2	Q59FF4_HUMAN	Q59FF4 homo sapien
38	9	2.3	3144	1	HD_HUMAN	P42858 homo sapien
39	8	2.0	54	2	Q6PPZ4_PLAFA	Q6PPZ4 plasmodium
40	8	2.0	67	2	Q18876_MACFA	Q18876 macaca fasc
41	8	2.0	131	2	Q6PPZ8_PLAFA	Q6PPZ8 plasmodium
42	8	2.0	134	2	Q4NX48_9DELT	Q4NX48 anaeromyxob
43	8	2.0	143	2	Q9FRF4_ORYSA	Q9FRF4 oryza sativ
44	8	2.0	158	2	Q762I8_ONCNE	Q762I8 oncorhynchu
45	8	2.0	174	2	Q8XVB3_RALSO	Q8XVB3 ralstonia s

ALIGNMENTS

RESULT 1

Q8K339_MOUSE PRELIMINARY; PRT; 391 AA.

AC Q8K339;

DT 01-OCT-2002 (TREMBlrel. 22, Created)

DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)

DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)

DE Antigenic determinant of rec-A protein.

GN Name=Kin;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;

OC Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;

RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,

RA Bosak S., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Vallalao D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,

RT "Generation and initial analysis of more than 15,000 full-length human

RL and mouse cDNA sequences."

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [2]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;

RA Strausberg R.;

RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

[3]

RP NUCLEOTIDE SEQUENCE.

RX MEDLINE=32020193; PubMed=1923796;

RA Angulo J., Rouer E., Mazin A., Mattei M., Tissier A., Horellou P.,

RA Benarous R., Devoret R.;

RT "Identification and expression of the cDNA of KIN17, a zinc-finger

RL protein."

RL Nucleic Acids Res. 19:5117-5123(1991).

DR EMBL; BC028860; AAH28860.1; -; mRNA.

DR PIR; S18666; S18666.


```
DR InterPro; IPR007087; Znf_C2H2_1.
DR SMART; SM00355; Znf_C2H2_1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
FT NON_TER 195
SQ SEQUENCE 195 AA; 23518 MW; 7E112E9CA820BFA6 CRC64;

Query Match 49.9%; Score 195; DB 2; Length 195;
Best Local Similarity 100.0%; Pred. No. 1.2e-188;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGKSDFLSPKAIANRIKSGKGLQKLRWYQCMQCKQCRDENGFKCHCMSESHQRLQLLASEN 60
DB 1 MGKSDFLSPKAIANRIKSGKGLQKLRWYQCMQCKQCRDENGFKCHCMSESHQRLQLLASEN 60
QY 61 PQQFMDYFSEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREHIHNNATQWETLTDFTK 120
DB 61 PQQFMDYFSEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREHIHNNATQWETLTDFTK 120
QY 121 WLGRGLCKVDPTKGYIQYIDRDPETIRROLELEKKKKQDLDDEKTAKFIEEQVRRG 180
DB 121 WLGRGLCKVDPTKGYIQYIDRDPETIRROLELEKKKKQDLDDEKTAKFIEEQVRRG 180
QY 181 LEGKEQETPVFTLS 195
DB 181 LEGKEQETPVFTLS 195

RESULT 3
ID O60870 HUMAN PRELIMINARY; PRT; 393 AA.
AC O60870;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Kin17 protein (HsKin17 protein) (KIN, antigenic determinant of recA
DE protein homolog) (MouSe).
GN Name:KIN; Synonym:s-Kin17; ORFNames=RP11-264C14.1-001;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Testis;
RX MEDLINE=20419742; PubMed=10964102; DOI=10.1093/carcin/21.9.1701;
RA Kannouche P., Mauffrey P., Pinon-Lataillade G., Mattei M.G.,
RA Sarasin A., Daya-Grosjean L., Angulo J.F.;
RT "Molecular cloning and characterization of the human KIN17 cDNA
RT encoding a component of the UV response that is conserved among
RL metazoans.";
RL Carcinogenesis 21:1701-1710(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Pancreas;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettner M., Madan A.A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
```

```
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Pancreas;
RA Strausberg R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP NUCLEOTIDE SEQUENCE.
RA Lovell J.; (MAY-2005) to the EMBL/GenBank/DBJ databases.
RL EMBL; AJ005273; CA006462.1; -; mRNA.
DR EMBL; BC017309; AAH17309.1; -; mRNA.
DR EMBL; AL158044; CA112959.1; -; Genomic DNA.
DR Ensembl; ENSG00000151657; Homo sapiens.
DR HGNC; HGNC:6327; KIN.
DR GO; GO:0005634; C:nucleus; TAS.
DR GO; GO:0003677; F:DNA binding; TAS.
DR InterPro; IPR005824; KOW.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00467; KOW; 1.
DR SMART; SM00355; Znf_C2H2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
KW Nuclear protein.
SQ SEQUENCE 393 AA; 45374 MW; 515A89B4C8A4C007 CRC64;

Query Match 45.8%; Score 179; DB 2; Length 393;
Best Local Similarity 100.0%; Pred. No. 3.7e-172;
Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 PKAIANRIKSGKGLQKLRWYQCMQCKQCRDENGFKCHCMSESHQRLQLLASENPQFMDYF 68
DB 9 PKAIANRIKSGKGLQKLRWYQCMQCKQCRDENGFKCHCMSESHQRLQLLASENPQFMDYF 68
QY 69 SESEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREHIHNNATQWETLTDFTKWLGRGLC 128
DB 69 SESEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREHIHNNATQWETLTDFTKWLGRGLC 128
QY 129 KVDETPTKGYIQYIDRDPETIRROLELEKKKKQDLDDEKTAKFIEEQVRRGLEKQEQE 187
DB 129 KVDETPTKGYIQYIDRDPETIRROLELEKKKKQDLDDEKTAKFIEEQVRRGLEKQEQE 187

RESULT 4
Q6GL97_XENTR PRELIMINARY; PRT; 389 AA.
AC Q6GL97;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE MGC69184 protein.
GN Name=MGC69184;
OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus; Silurana.
OX NCBI_TaxID=8364;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Embryo;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
```

RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skaleka U., Smailus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Embryo;
 RA Klein S., Gerhard D.S.;
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC074606; AAH74606.1; -; mRNA.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
 DR PROSITE; PS00028; ZINC FINGER C2H2 1; UNKNOWN 1.
 SQ SEQUENCE 389 AA; 44387 MW; 793E70C997451P6A CRC64;

 Query Match 15.6%; Score 61; DB 2; Length 389;
 Best Local Similarity 100.0%; Pred. No. 1.7e-52;
 Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 9 PKAIANRIKSGKGLRWYCOMQKQCRDENGFKCHCMSESHQRLLLASENPQQFMDYF 68
 DB 9 PKAIANRIKSGKGLRWYCOMQKQCRDENGFKCHCMSESHQRLLLASENPQQFMDYF 68

 QY 69 S 69
 DB 69 S 69

 RESULT 5
 Q68F56_XENLA
 ID Q68F56_XENLA PRELIMINARY; PRT; 387 AA.
 AC Q68F56;
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DE 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE MGC81626 protein.
 GN Name=MGC81626;
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
 OC Xenopodinae; Xenopus; Xenopus.
 OX NCBI_TaxID=8355;
 OX [1]
 RN NUCLEOTIDE SEQUENCE.
 RP TISSUE=Embryo;
 RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
 RA Richardson P.;
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
 initiative.";
 RL Dev. Dyn. 225:384-391 (2002).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Embryo;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Uddin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skaleka U., Smailus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Embryo;
 RA Klein S., Gerhard D.S.;
 RL Submitted (AUG-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC079988; AAH79988.1; -; mRNA.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0005929; C:ribonucleoprotein complex; IEA.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR InterPro; IPR005824; KOW.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00467; KOW; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2 1; 1.
 SQ SEQUENCE 387 AA; 44491 MW; 4CCEB56B047F3779 CRC64;

 Query Match 12.8%; Score 50; DB 2; Length 387;
 Best Local Similarity 100.0%; Pred. No. 2.4e-41;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 14 NRIKSGKGLRWYCOMQKQCRDENGFKCHCMSESHQRLLLASENPQQ 63
 DB 14 NRIKSGKGLRWYCOMQKQCRDENGFKCHCMSESHQRLLLASENPQQ 63

 RESULT 6
 Q455G2_TETNG
 ID Q455G2_TETNG PRELIMINARY; PRT; 427 AA.
 AC Q455G2;
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
 DE Chromosome 19 SCAP14731, whole genome shotgun sequence.
 GN ORFNames=GSTENG00023750001;
 OS Tetraodon nigroviridis (Green puffer).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 OC Tetraodontidae; Tetraodontidae; Tetraodon.
 OX NCBI_TaxID=99883;
 OX [1]
 RN NUCLEOTIDE SEQUENCE.
 RP Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
 RA Mauceli E., Bounneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
 RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dessat C., Segurens B.,
 RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
 RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
 RA Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
 RA Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
 RA Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
 RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
 RA Wincker P., Lander E.S., Weissenbach J., Roest Crollius H.;
 RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
 the early vertebrate proto-karyotype.";
 RL Nature 431:946-957 (2004).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RG Genoscope; Whitehead Institute Centre for Genome Research;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 preliminary data.
 DR EMBL; CAAE01014731; CAG04120.1; -; Genomic DNA.
 SQ SEQUENCE 427 AA; 48007 MW; ADF91299B8CA23B CRC64;

Query Match 12.3%; Score 48; DB 2; Length 427;
 Best Local Similarity 100.0%; Pred. No. 2.8e-39; Mismatches 0; Indels 0; Gaps 0;
 Matches 48; Conservative 0;

QY 14 NRIKSGQLRWYQCMQKQCRDENGFKCHMSHQRQLLASNP 61
 DB 14 NRIKSGQLRWYQCMQKQCRDENGFKCHMSHQRQLLASNP 61
 |||||

RESULT 7
 Q5RH25 BRARE PRELIMINARY; PRT; 383 AA.
 AC Q5RH25
 DT 01-FEB-2005 (TrEMBLrel. 29, Created)
 DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
 DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
 DE Novel protein similar to vertebrate KIN, antigenic determinant of reca
 DE protein homolog (Mouse) (KIN).
 GN Name=OTTDARP0000008158; ORFNames=CH211-235A22.3-001;
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Hunter G.;
 RL Submitted (DEC-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BX640594; CA11855.1; -; Genomic_DNA.
 DR InterPro; IPR005824; KOW.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00467; KOW; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
 SQ SEQUENCE 383 AA; 44337 MW; 1F0AA34FB62176D1 CRC64;

Query Match 11.0%; Score 43; DB 2; Length 383;
 Best Local Similarity 100.0%; Pred. No. 3e-34; Mismatches 0; Indels 0; Gaps 0;
 Matches 43; Conservative 0;

QY 19 KGLQKLWYQCMQKQCRDENGFKCHMSHQRQLLASNP 61
 DB 19 KGLQKLWYQCMQKQCRDENGFKCHMSHQRQLLASNP 61
 |||||

RESULT 8
 Q4TDV9 TETNG PRELIMINARY; PRT; 269 AA.
 AC Q4TDV9
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
 DE Chromosome undetermined SCAF6005, whole genome shotgun sequence.
 DE (Fragment).
 GN ORFNames=GSTENG0002601001;
 OS Tetraodon nigroviridis (Green puffer).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 OC Tetraodontidae; Tetraodontidae; Tetraodon.
 OX NCBI_TaxID=99883;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Jallou O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
 RA Mauceli E., Bounneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
 RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segreus B.,
 RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
 RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
 RA Bismont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
 RA Craud C., Duprat S., Brottier P., Coutanceau J.P., Guzy J.,
 RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
 RA Kellis M., Wolff J.N., Guigo R., Zody M.C., Mesirov J.,
 RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,

RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
 RA Wincker P., Lander E.S., Weissenbach J., Roest Crolius H.;
 RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
 RT the early vertebrate proto-karyotype.";
 RL Nature 431:946-957(2004).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RG Genoscope; Whitehead Institute Centre for Genome Research;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 CC -1- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 CC EMBL; CAE01006005; CAF88923.1; -; Genomic_DNA.
 FT NON_TER 1 1
 FT TER 269 269
 SQ SEQUENCE 269 AA; 31736 MW; E6883839A2CD17AF CRC64;

Query Match 10.5%; Score 41; DB 2; Length 269;
 Best Local Similarity 100.0%; Pred. No. 2.3e-32; Mismatches 0; Indels 0; Gaps 0;
 Matches 41; Conservative 0;

QY 112 WETLDTFTKWLGRGLCKVDTEPKGWYQYIDRDPETIRQ 152
 DB 58 WETLDTFTKWLGRGLCKVDTEPKGWYQYIDRDPETIRQ 98
 |||||

RESULT 9
 Q9AY87 ORYSA PRELIMINARY; PRT; 402 AA.
 AC Q9AY87
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Hypothetical protein OSJNBa0004B24.17.
 GN Names=OSJNBa0004B24.17;
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Burr P.C., Hsiao J.,
 RA Zismann V., Pai G., Bowman C.L., Fujii C.Y., VanAken S.E.,
 RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblyum T.V.,
 RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RA Buell R.;
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC084319; AAC59654.1; -; Genomic_DNA.
 DR Gramene; Q9AY87; -.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0046872; F:metal ion binding; IEA.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR SMART; IPR007087; Znf_C2H2.
 DR SMART; SM00355; Znf_C2H2; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
 KW Hypothetical protein.
 SQ SEQUENCE 402 AA; 45962 MW; A1F4E38ECC60D13E CRC64;

Query Match 9.0%; Score 35; DB 2; Length 402;
 Best Local Similarity 100.0%; Pred. No. 4e-26; Mismatches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLWYQCMQKQCRDENGFKCHMSHQRQ 53
 DB 19 KGLQKLWYQCMQKQCRDENGFKCHMSHQRQ 53
 |||||

RESULT 10

O9ZVU5 ARATH
 ID Q9ZVU5_ARATH PRELIMINARY; PRT; 411 AA.
 AC Q9ZVU5;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
 DE T5A14.13 protein (Hypothetical protein At1955460).
 GN Name=T5A14.13; Synonyms=At1955460;
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eusoids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
 RA Altafi H., Araujo R., Huizar L., Rowley D., Buehler E., Dunn P.,
 RA Gonzalez A., Kremenetskaia I., Kim C., Lenz C., Li J., Liu S.,
 RA Lurss S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,
 RA Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Etgu P., Lee J.M.,
 RA Toriumi M., Yu G., Brooks S., Chao Q., Chen H., Karlin-Neumann G.,
 RA Kim C., Lam B., Miranda M., Nguyen M., Palm C.J., Shinn P.,
 RA Southwick A., Davis R.W., Ecker J.R., Theologis A.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 RA Goldsmith A.D., Lee J.M., Quach H.B., Toriumi M., Yu G., Bowser L.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
 RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC005223; AAD10649.1; -; Genomic_DNA.
 DR EMBL; AY051011; AAK93688.1; -; mRNA.
 DR EMBL; AF360132; AAK25842.1; -; mRNA.
 DR PIR; H96596; H96596.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR InterPro; IPR005824; KOW.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00467; KOW; 1.
 DR SMART; SM00739; KOW; 1.
 DR SMART; SM00355; Znf_C2H2; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
 KW Hypothetical protein.
 SQ SEQUENCE 411 AA; 47288 MW; 9DA6F8648002065D CRC64;
 Query Match 9.0%; Score 35; DB 2; Length 411;
 Best Local Similarity 100.0%; Pred. No. 4.1e-26;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 19 KGLQKLRYWCQCKQCRDENGFKCHCMSESHQRQ 53
 Db |||
 19 KGLQKLRYWCQCKQCRDENGFKCHCMSESHQRQ 53
 RESULT 11
 Q75LU5 ORYSA
 ID Q75LU5_ORYSA PRELIMINARY; PRT; 430 AA.
 AC Q75LU5;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Hypothetical protein OSJNB0015121.3.
 GN Name=OSJNB0015121.3;

OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,
 RA Overton II L.B., Tsitrin T., Kim M.M., Bera J.J., Jin S.S.,
 RA Padrosh D.W., Tallon L.J., Koo H., Ziemann V., Hsiao J., Blunt S.,
 RA Vanaken S.S., Riedmuller S.B., Utterback T.T., Feidblyum T.V.,
 RA Yang Q.Q., Haas B.J., Suh B.B., Peterson J.J., Quackenbush J.,
 RA White O., Salzberg S.L., Fraser C.M.;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RA Buell R.;
 RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC091302; AAR00634.1; -; Genomic_DNA.
 DR Gramene; Q75LU5; -;
 DR InterPro; IPR005824; KOW.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00467; KOW; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
 KW Hypothetical protein.
 SQ SEQUENCE 430 AA; 49067 MW; 6151D748ED2F4690 CRC64;
 Query Match 9.0%; Score 35; DB 2; Length 430;
 Best Local Similarity 100.0%; Pred. No. 4.3e-26;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 19 KGLQKLRYWCQCKQCRDENGFKCHCMSESHQRQ 53
 Db |||
 19 KGLQKLRYWCQCKQCRDENGFKCHCMSESHQRQ 53
 RESULT 12
 Q4TDMO_TETNG
 ID Q4TDMO_TETNG PRELIMINARY; PRT; 88 AA.
 AC Q4TDMO;
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
 DE Chromosome undetermined SCAF6004, whole genome shotgun sequence.
 DE (Fragment).
 GN ORFNames=GSTENG000260001;
 OS Tetraodon nigroviridis (Green puffer).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 OC Tetraodontidae; Tetraodontidae; Tetraodon.
 OX NCBI_TaxID=99883;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
 RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
 RA Nicod S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
 RA Dasilva C., Salancoubat M., Levy M., Boudet N., Castellano S.,
 RA Anthonard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
 RA Bismont C., Skalli Z., Cattolico L., Poulin J., De Berardinis V.,
 RA Craud C., Duprat S., Brottier P., Coutanceau J.P., Guzy J.,
 RA Parra G., Lardier G., Chapelle C., McKernan K.J., McEwan P., Bosak S.,
 RA Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
 RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
 RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
 RA Wincker P., Lander E.S., Weissbach J., Roest Crolius H.;
 RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
 the early vertebrate proto-karyotype.";
 RL Nature 431:946-957(2004).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RG Genoscope; Whitehead Institute Centre for Genome Research;
 RP Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.

CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data!
DR EMBL; CAAE01006004; CAF88922.1; -; Genomic_DNA.
FT NON_TER 1 1
FT NON_TER 88 88
SQ SEQUENCE 88 AA; 9709 MW; C5BD169AF081CBC2 CRC64;

Query Match 5.1%; Score 20; DB 2; Length 88;
Best Local Similarity 100.0%; Pred. No. 1.7e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 323 LKLDQTHLEIVIPAGKRVL 342
Db 21 LKLDQTHLEIVIPAGKRVL 40
|||||

RESULT 13
O6XI90_DROYA
ID Q6XI90_DROYA PRELIMINARY; PRT; 178 AA.
AC Q6XI90;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Similar to Drosophila melanogaster kin17 (Fragment).
OS Drosophila yakuba (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7245;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22887302; PubMed=14525923; DOI=10.1101/gr.1311003;
RA Domazet-Lozo T., Tautz D.;
RT "An evolutionary analysis of orphan genes in Drosophila.";
RL Genome Res. 13:2213-2219(2003).
DR EMBL; AY231941; AAR09964.1; -; mRNA.
DR InterPro; IPR007087; Znf_C2H2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
FT NON_TER 1 1
FT NON_TER 178 178
SQ SEQUENCE 178 AA; 21188 MW; B251B8630CBC76C3 CRC64;

Query Match 3.8%; Score 15; DB 2; Length 178;
Best Local Similarity 100.0%; Pred. No. 3.7e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 KSKGLQKLRYCQMC 31
Db 17 KSKGLQKLRYCQMC 31
|||||

RESULT 14
O76926_DROME
ID O76926_DROME PRELIMINARY; PRT; 244 AA.
AC O76926;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE KIN17 protein (Fragment).
GN Name=kin17; ORFNames=CG5649;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=EST28;
RA Garcia V., de La Roche Saint Andre C.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ006528; CAA07089.1; -; Genomic_DNA.
DR Ensembl; CG5649; Drosophila melanogaster.

DR FlyBase; FBgn0024887; CG5649.
DR FlyBase; FBgn0024887; kin17.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR007087; Znf_C2H2.
DR SMART; SM00355; Znf_C2H2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
FT NON_TER 244 244
SQ SEQUENCE 244 AA; 28679 MW; 5AF6C28524234582 CRC64;

Query Match 3.8%; Score 15; DB 2; Length 244;
Best Local Similarity 100.0%; Pred. No. 4.9e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 KSKGLQKLRYCQMC 31
Db 17 KSKGLQKLRYCQMC 31
|||||

RESULT 15
Q8XR2_DROME
ID Q8XR2_DROME PRELIMINARY; PRT; 390 AA.
AC Q8XR2;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE R655257P.
GN Name=kin17; ORFNames=CG5649;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Murgall C.J., Nunco J., Pacleb J., Parages V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY084184; AAL89922.1; -; mRNA.
DR FlyBase; FBgn0024887; CG5649.
DR FlyBase; FBgn0024887; kin17.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR007087; Znf_C2H2.
DR SMART; SM00355; Znf_C2H2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
KW Nuclear protein.
SQ SEQUENCE 390 AA; 45286 MW; 39590CD3B61077CC CRC64;

Query Match 3.8%; Score 15; DB 2; Length 390;
Best Local Similarity 100.0%; Pred. No. 7.5e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 KSKGLQKLRYCQMC 31
Db 17 KSKGLQKLRYCQMC 31
|||||

Search completed: November 23, 2005, 16:49:37
Job time : 155.106 secs

This Page Blank (usps)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:35:49 ; Search time 32.9158 Seconds
(without alignments)
982.087 Million cell updates/sec

Title: US-09-555-529-25
Perfect score: 391
Sequence: 1 MCKSDFLSPKAIANRIKSG.....GPLKGRVGIQVEDISKLA 391

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 572060 seqs, 82675679 residues

Word size : 0

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2.6/prodata/1/iaa/5 COMB.pcp.*
2: /cgn2.6/prodata/1/iaa/6 COMB.pcp.*
3: /cgn2.6/prodata/1/iaa/H COMB.pcp.*
4: /cgn2.6/prodata/1/iaa/PTUS COMB.pcp.*
5: /cgn2.6/prodata/1/iaa/RE COMB.pcp.*
6: /cgn2.6/prodata/1/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	2.3	1543	2	US-09-904-987-7
2	9	2.3	3144	1	US-08-246-982A-6
3	9	2.3	3144	1	US-08-453-265-6
4	9	2.3	3144	1	US-08-457-273B-42
5	9	2.3	3144	2	US-08-556-419-21
6	9	2.3	3144	2	US-09-041-886-15
7	9	2.3	3144	2	US-09-538-092-1118
8	8	2.0	136	2	US-09-270-767-32008
9	8	2.0	400	2	US-09-826-509-491
10	8	2.0	444	1	US-07-626-618A-20
11	8	2.0	444	1	US-08-333-977-20
12	8	2.0	446	1	US-07-781-254A-2
13	8	2.0	446	1	US-07-781-254A-3
14	8	2.0	546	2	US-09-252-991A-18637
15	8	2.0	562	2	US-09-949-002-426
16	8	2.0	937	1	US-08-253-155A-31
17	8	2.0	937	2	US-09-538-092-1092
18	8	2.0	937	2	US-09-949-002-374
19	8	2.0	959	2	US-09-538-092-1091
20	7	1.8	58	2	US-07-741-453A-41
21	7	1.8	65	2	US-09-881-572A-1
22	7	1.8	68	2	US-09-270-767-34723
23	7	1.8	68	2	US-09-270-767-49940
24	7	1.8	74	2	US-09-489-039A-7206
25	7	1.8	76	2	US-09-621-976-7070
26	7	1.8	80	2	US-08-930-503A-16
27	7	1.8	90	2	US-09-613-486-24

28	7	1.8	91	2	US-09-134-001C-4276	Sequence 4276, Ap
29	7	1.8	98	2	US-09-489-039A-7316	Sequence 7316, Ap
30	7	1.8	99	2	US-09-866-108A-15755	Sequence 15755, A
31	7	1.8	120	2	US-09-311-021-20	Sequence 20, Appl
32	7	1.8	152	2	US-09-248-796A-15608	Sequence 15608, A
33	7	1.8	197	2	US-09-489-039A-14336	Sequence 14336, A
34	7	1.8	201	2	US-09-744-778-4	Sequence 4, Appl
35	7	1.8	208	2	US-09-252-991A-32430	Sequence 32430, A
36	7	1.8	218	2	US-09-949-016-6586	Sequence 6586, Ap
37	7	1.8	229	2	US-09-904-923C-1	Sequence 1, Appl
38	7	1.8	238	1	US-08-103-170-13	Sequence 13, Appl
39	7	1.8	250	2	US-09-809-665A-83	Sequence 83, Appl
40	7	1.8	253	2	US-09-248-796A-15042	Sequence 15042, A
41	7	1.8	262	2	US-09-489-039A-12195	Sequence 12195, A
42	7	1.8	263	2	US-09-902-540-10610	Sequence 10610, A
43	7	1.8	271	2	US-09-949-016-11720	Sequence 11720, A
44	7	1.8	277	1	US-08-400-413-1	Sequence 1, Appl
45	7	1.8	299	2	US-09-248-796A-15986	Sequence 15986, A

ALIGNMENTS

RESULT 1
US-09-904-987-7
; Sequence 7, Application US/09904987
; Patent No. 6743771
; GENERAL INFORMATION:
; APPLICANT: No. 6743771actyl, Inc.
; TITLE OF INVENTION: Methods and Compositions for Controlling Pathological and Prepatk
; FILE OF INVENTION: Protein Assembly or Aggregation
; FILE REFERENCE: 42108/26146
; CURRENT APPLICATION NUMBER: US/09/904, 987
; CURRENT FILING DATE: 2001-07-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 7
; LENGTH: 1543
; TYPE: PRT
; ORGANISM: homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NCBI ENTREZ / XP_003405
; DATABASE ENTRY DATE: 2001-04-16
; RELEVANT RESIDUES: (1)..(1543)
US-09-904-987-7

Query Match 2.3%; Score 9; DB 2; Length 1543;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 218 TTSKSSSLG 226
Db 1219 TTSKSSSLG 1227

RESULT 2
US-08-246-982A-6
; Sequence 6, Application US/08246982A
; Patent No. 5686288
; GENERAL INFORMATION:
; APPLICANT: MacDonald, Marcy E.
; APPLICANT: Ambrose, Christine M.
; APPLICANT: Guyao, Mabel P.
; APPLICANT: Guebla, James F.
; TITLE OF INVENTION: Huntingtin DNA, Protein And Uses Thereof
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/246,982A
FILING DATE: May 20, 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Goldstein, Jorge, A.
REGISTRATION NUMBER: 29,021
REFERENCE/DOCKET NUMBER: 0609.3880002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3144 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-246-982A-6

Query Match 2.3%; Score 9; DB 1; Length 3144;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 TTSKSSSLG 226
DB 1222 TTSKSSSLG 1230

RESULT 3
US-08-453-265-6
Sequence 6, Application US/08453265
Patent No. 5635757
GENERAL INFORMATION:
APPLICANT: Macdonald, Marcy E.
APPLICANT: Ambrose, Christine M.
APPLICANT: Duyao, Mabel P.
APPLICANT: Guehlla, James P.
TITLE OF INVENTION: Huntingtin DNA, Protein And Uses Thereof
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSES: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/453,265
FILING DATE: 30-MAY-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Ludwig, Steven R.
REGISTRATION NUMBER: 36,203
REFERENCE/DOCKET NUMBER: 0609.3880003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3144 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-453-265-6
Query Match 2.3%; Score 9; DB 1; Length 3144;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 TTSKSSSLG 226
DB 1222 TTSKSSSLG 1230

RESULT 4
US-08-457-273B-42
Sequence 42, Application US/08457273B
Patent No. 5849995
GENERAL INFORMATION:
APPLICANT: Hayden, Michael
APPLICANT: Lin, Biaoyang
APPLICANT: Nasir, Jamal
TITLE OF INVENTION: Mouse Model for Huntington's Disease and
TITLE OF INVENTION: Related DNA Sequences
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSES: Virginia Bennett
STREET: PO Box 37428
CITY: Raleigh
STATE: No. 5849995th Carolina
COUNTRY: US
ZIP: 27627
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/457,273B
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Bennett, Virginia C.
REGISTRATION NUMBER: 37,092
REFERENCE/DOCKET NUMBER: 3477-85A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-854-1400
TELEFAX: 919-854-1401
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 3144 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-457-273B-42

Query Match 2.3%; Score 9; DB 1; Length 3144;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 TTSKSSSLG 226
DB 1222 TTSKSSSLG 1230

RESULT 5
US-08-556-419-21
Sequence 21, Application US/08556419C
Patent No. 6093549
GENERAL INFORMATION:
APPLICANT: Ross, Christopher
APPLICANT: Li, Xiao-Jiang
APPLICANT: Li, Shi-Hua
APPLICANT: Sharp, Alan
APPLICANT: Lanahan, Anthony

APPLICANT: Worley, Paul
APPLICANT: Snyder, Solomon
TITLE OF INVENTION: Huntington-associated protein
FILE REFERENCE: 01107.52271
CURRENT APPLICATION NUMBER: US/08/556,419C
CURRENT FILING DATE: 1995-11-09
NUMBER OF SEQ ID NOS: 25
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 21
LENGTH: 3144
TYPE: PRT
ORGANISM: Homo sapiens
US-08-556-419-21

Query Match 2.3%; Score 9; DB 2; Length 3144;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 218 TTSKSSSLG 226
Db 1222 TTSKSSSLG 1230

RESULT 6
US-09-041-886-15
Sequence 15, Application US/09041886
Patent No. 6235872
GENERAL INFORMATION:
APPLICANT: Bredesen, Dale E.
APPLICANT: Razizadeh, Sharoz
TITLE OF INVENTION: Protoprotic Peptides, Dependence
TITLE OF INVENTION: Polypeptides and Methods of Use
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/041,886
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 2626
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 3144 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-041-886-15

Query Match 2.3%; Score 9; DB 2; Length 3144;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 218 TTSKSSSLG 226
Db 1222 TTSKSSSLG 1230

RESULT 7
US-09-538-092-1118
Sequence 1118, Application US/09538092
Patent No. 6753314
GENERAL INFORMATION:
APPLICANT: Giot, Loic
APPLICANT: Mansfield, Traci A.
TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
FILE REFERENCE: 15966-542
CURRENT APPLICATION NUMBER: US/09/538,092
CURRENT FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: 60/127,352
PRIOR FILING DATE: 1999-04-01
PRIOR APPLICATION NUMBER: 60/178,965
PRIOR FILING DATE: 2000-02-01
NUMBER OF SEQ ID NOS: 1387
SOFTWARE: CurapatSeqformat Version 0.9
SEQ ID NO 1118
LENGTH: 3144
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (0)...(0)
OTHER INFORMATION: Polypeptide Accession Number P42858
US-09-538-092-1118

Query Match 2.3%; Score 9; DB 2; Length 3144;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 218 TTSKSSSLG 226
Db 1222 TTSKSSSLG 1230

RESULT 8
US-09-270-767-32008
Sequence 32008, Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:
APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 32008
LENGTH: 136
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-09-270-767-32008

Query Match 2.0%; Score 8; DB 2; Length 136;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 329 HLETVIPA 336
Db 72 HLETVIPA 79

RESULT 9
US-09-826-509-491
Sequence 491, Application US/09826509
Patent No. 6806054
GENERAL INFORMATION:
APPLICANT: Lehmann-Bruinsma, Karin
APPLICANT: Liaw, Chen W.
APPLICANT: Lin, I-Lin
TITLE OF INVENTION: No. 6806054-Endogenous, Constitutively Activated Known G
TITLE OF INVENTION: Protein-Coupled Receptors

FILE REFERENCE: AREN-207
CURRENT APPLICATION NUMBER: US/09/826,509
PRIOR FILING DATE: 2001-04-05
PRIOR APPLICATION NUMBER: 60/195,747
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: 09/170,496
PRIOR FILING DATE: 1998-10-13
NUMBER OF SEQ ID NOS: 589
SOFTWARE: Patentin Version 2.1
SEQ ID NO 491
LENGTH: 400
TYPE: PRT
ORGANISM: Homo sapiens
US-09-826-509-491

Query Match 2.0%; Score 8; DB 2; Length 400;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 IDRYTAVV 314
Db 126 IDRYTAVV 133

RESULT 10
US-07-626-618A-20
Sequence 20, Application US/07626618A
Patent No. 5422265
GENERAL INFORMATION:
APPLICANT: Van Tol, Hubert H.M.
APPLICANT: Civeilli, Olivier
TITLE OF INVENTION: A No. 5422265el Human Dopamine Receptor and Uses
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Allegretti & Witcoff, Ltd.
STREET: 10 South Wacker Drive, Suite 3000
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/626,618A
FILING DATE: 7 DEC 1990
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: NO. 5422265nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 90,1092
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-715-1000
TELEFAX: 312-715-1234
TELEX: 810-221-8317
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 444 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO

Query Match 2.0%; Score 8; DB 1; Length 444;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 IDRYTAVV 314
Db 126 IDRYTAVV 133

Query Match 2.0%; Score 8; DB 1; Length 444;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 IDRYTAVV 314
Db 126 IDRYTAVV 133

RESULT 11
US-08-333-977-20
Sequence 20, Application US/08333977
Patent No. 5594108
GENERAL INFORMATION:
APPLICANT: Van Tol, Hubert H.M.
APPLICANT: Civeilli, Olivier
TITLE OF INVENTION: A No. 5594108el Human Dopamine Receptor and Uses
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Allegretti & Witcoff, Ltd.
STREET: 10 South Wacker Drive, Suite 3000
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/333,977
FILING DATE: 03-NOV-1994
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/626,618
FILING DATE: 7 DEC 1990
ATTORNEY/AGENT INFORMATION:
NAME: NO. 5594108nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 90,1092
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-715-1000
TELEFAX: 312-715-1234
TELEX: 810-221-8317
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 444 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO

Query Match 2.0%; Score 8; DB 1; Length 444;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 IDRYTAVV 314
Db 126 IDRYTAVV 133

RESULT 12
US-07-781-254A-2
Sequence 2, Application US/07781254A
Patent No. 5407823
GENERAL INFORMATION:
APPLICANT: Sokoloff, Pierre
APPLICANT: Martress, Marie-Pascale
APPLICANT: Schwartz, Jean-Charles
APPLICANT: Giros, Bruno
TITLE OF INVENTION: Polypeptides Having a Dopaminergic Receptor
TITLE OF INVENTION: Activity, Nucleic Acids Coding for These Polypeptides and the
TITLE OF INVENTION: Use of These Polypeptides for the Screening of Substances
TITLE OF INVENTION: Active on These Polypeptides
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 3100 No. 5407823west Center

;; CITY: Minneapolis
;; STATE: MN
;; COUNTRY: USA
;; ZIP: 55402
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/07/781,254A
;; FILING DATE: 31-DEC-1991
;; CLASSIFICATION: 530
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Hillson, Randall A.
;; REGISTRATION NUMBER: 31,939
;; REFERENCE/DOCKET NUMBER: 8076.51-WOUS
;;
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 612-332-5300
;; TELEFAX: 612-332-9081
;;
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 446 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; IMMEDIATE SOURCE:
;; CLONE: Variant D-3 dopaminergic receptor
;;
US-07-781-254A-2

Query Match 2.0%; Score 8; DB 1; Length 446;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 IDRYTAVV 314
Db 126 IDRYTAVV 133
|||||

RESULT 13
US-07-781-254A-3
; Sequence 3, Application US/07781254A
; Patent No. 5407823
; GENERAL INFORMATION:
; APPLICANT: Sokoloff, Pierre
; APPLICANT: Martress, Marie-Pascale
; APPLICANT: Schwartz, Jean-Charles
; APPLICANT: Giros, Bruno
; TITLE OF INVENTION: Polypeptides Having a Dopaminergic Receptor
; TITLE OF INVENTION: Activity, Nucleic Acids Coding for These Polypeptides and the
; TITLE OF INVENTION: Use of These Polypeptides for the Screening of Substances
; TITLE OF INVENTION: Active on These Polypeptides
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 3100 No. 5407823west Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/781,254A
; FILING DATE: 31-DEC-1991
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Hillson, Randall A.
; REGISTRATION NUMBER: 31,939
; REFERENCE/DOCKET NUMBER: 8076.51-WOUS

;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 612-332-5300
;; TELEFAX: 612-332-9081
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 446 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; IMMEDIATE SOURCE:
;; CLONE: D-3 dopaminergic receptor
;;
US-07-781-254A-3

Query Match 2.0%; Score 8; DB 1; Length 446;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 IDRYTAVV 314
Db 126 IDRYTAVV 133
|||||

RESULT 14
US-09-252-991A-18637
; Sequence 18637, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18637
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18637

Query Match 2.0%; Score 8; DB 2; Length 546;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 GAGGSAGA 217
Db 520 GAGGSAGA 527
|||||

RESULT 15
US-09-949-002-426
; Sequence 426, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: PatsSeq for Windows Version 4.0
; SEQ ID NO 426
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Human

US-09-949-002-426

Query Match 2.0%; Score 8; DB 2; Length 562;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 GAGGSAGA 217
|||
Db 96 GAGGSAGA 103

Search completed: November 23, 2005, 16:50:48
Job time : 33.9158 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:44:35 ; Search time 110.717 Seconds
(without alignments)
1475.577 Million cell updates/sec

Title: US-09-555-529-25
Perfect score: 391
Sequence: 1 MGKSDPFSKAIANRIKSG.....GPLKGRVEGIQVEDISKLA 391

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1867569 seqs, 417829326 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA Main:
1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pap:*
2: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pap:*
3: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pap:*
4: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pap:*
5: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pap:*
6: /cgn2_6/prodata/1/pubpaa/US11_PUBCOMB.pap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	179	45.8	393	4	US-10-408-765A-1465
2	35	9.0	120	4	US-10-767-701-57151
3	35	9.0	138	4	US-10-437-963-191916
4	35	9.0	423	5	US-10-739-930-8499
5	35	9.0	424	4	US-10-425-115-341788
6	35	9.0	437	4	US-10-425-114-48183
7	35	9.0	437	4	US-10-425-114-58982
8	30	7.7	176	4	US-10-424-599-168326
9	29	7.4	84	4	US-10-106-698-5464
10	15	3.8	390	6	US-11-097-143-41040
11	9	2.3	335	5	US-10-732-923-11274
12	9	2.3	535	5	US-10-739-930-10661
13	9	2.3	539	5	US-10-732-923-11269
14	9	2.3	1543	3	US-09-904-987-7
15	9	2.3	3141	4	US-10-408-765A-2250
16	9	2.3	3144	4	US-10-419-997-1
17	9	2.3	3223	4	US-10-276-774-1777
18	9	2.3	3223	4	US-10-276-774-1840
19	8	2.0	8	4	US-10-022-066-38
20	8	2.0	8	4	US-10-022-066-610
21	8	2.0	10	3	US-09-572-404B-3236
22	8	2.0	53	4	US-10-425-115-296022
23	8	2.0	143	4	US-10-437-963-176007
24	8	2.0	164	4	US-10-425-115-287597
25	8	2.0	183	6	US-11-097-143-27039
26	8	2.0	192	6	US-11-097-143-2925
27	8	2.0	198	3	US-09-794-257-16

28	8	2.0	198	3	US-09-945-173-5	Sequence 5, Appl1
29	8	2.0	198	3	US-09-972-529-4	Sequence 4, Appl1
30	8	2.0	198	4	US-10-400-991-70	Sequence 70, Appl1
31	8	2.0	222	4	US-10-180-375-110	Sequence 110, App
32	8	2.0	222	4	US-10-183-687-126	Sequence 126, App
33	8	2.0	222	4	US-10-424-599-202515	Sequence 202515, App
34	8	2.0	224	5	US-10-739-930-9455	Sequence 9455, App
35	8	2.0	229	4	US-10-180-375-112	Sequence 112, App
36	8	2.0	229	4	US-10-183-687-128	Sequence 128, App
37	8	2.0	229	4	US-10-424-599-178340	Sequence 178340, App
38	8	2.0	362	4	US-10-424-599-247365	Sequence 247365, App
39	8	2.0	400	3	US-09-826-509-431	Sequence 491, App
40	8	2.0	400	4	US-10-225-567A-104	Sequence 104, App
41	8	2.0	400	4	US-10-292-798-1244	Sequence 1244, App
42	8	2.0	400	5	US-10-925-095-491	Sequence 491, App
43	8	2.0	438	4	US-10-425-115-214055	Sequence 214055, App
44	8	2.0	461	4	US-10-282-122A-64750	Sequence 64750, A
45	8	2.0	494	4	US-10-282-122A-52864	Sequence 52864, A

ALIGNMENTS

RESULT 1
US-10-408-765A-1465
; Sequence 1465, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1465
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1465

Query Match	45.8%	Score 179;	DB 4;	Length 393;
Best Local Similarity	100.0%;	Pred. No. 4.4e-159;		
Matches 179;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	9	PKAIANRIKSGKLGKLRWTCQKCRDENGPKCHCMESHQROILLASENPQQFMDYF	68	
DB	9	PKAIANRIKSGKLGKLRWTCQKCRDENGPKCHCMESHQROILLASENPQQFMDYF	68	
QY	69	SEEFNDFLELLRRRRTKRVHNNIVYNEVISHREIHNNATOWETLDTFTKWLGRGLC	128	
DB	69	SEEFNDFLELLRRRRTKRVHNNIVYNEVISHREIHNNATOWETLDTFTKWLGRGLC	128	
QY	129	KVDETPKGWIQIYDRDPETIRRQLELEKKKKQDDEEKTAKFIEQVRRGLEKQE	187	
DB	129	KVDETPKGWIQIYDRDPETIRRQLELEKKKKQDDEEKTAKFIEQVRRGLEKQE	187	

RESULT 2
US-10-767-701-57151
; Sequence 57151, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53353)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 57151
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 30563818.pcp
US-10-767-701-57151

Query Match 9.0%; Score 35; DB 4; Length 120;
Best Local Similarity 100.0%; Pred. No. 2.4e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRLWYCMQCKQCRDENGFKCHCMSESHQRQ 53
|||
DB 19 KGLQKLRLWYCMQCKQCRDENGFKCHCMSESHQRQ 53

RESULT 3

US-10-437-963-191916
; Sequence 191916, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 191916
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_88190C.1.pcp
US-10-437-963-191916

Query Match 9.0%; Score 35; DB 4; Length 138;
Best Local Similarity 100.0%; Pred. No. 2.8e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRLWYCMQCKQCRDENGFKCHCMSESHQRQ 53
|||
DB 19 KGLQKLRLWYCMQCKQCRDENGFKCHCMSESHQRQ 53

RESULT 4

US-10-739-930-8499
; Sequence 8499, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 8499
; LENGTH: 423

; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: ZEAMA-23APR03-C7583_2.p
US-10-739-930-8499

Query Match 9.0%; Score 35; DB 5; Length 423;
Best Local Similarity 100.0%; Pred. No. 7.4e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRLWYCMQCKQCRDENGFKCHCMSESHQRQ 53
|||
DB 19 KGLQKLRLWYCMQCKQCRDENGFKCHCMSESHQRQ 53

RESULT 5

US-10-425-115-341788
; Sequence 341788, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 341788
; LENGTH: 424
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_74879C.1.pcp
US-10-425-115-341788

Query Match 9.0%; Score 35; DB 4; Length 424;
Best Local Similarity 100.0%; Pred. No. 7.4e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRLWYCMQCKQCRDENGFKCHCMSESHQRQ 53
|||
DB 19 KGLQKLRLWYCMQCKQCRDENGFKCHCMSESHQRQ 53

RESULT 6

US-10-425-114-48183
; Sequence 48183, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53131)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 48183
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3591-011-E3_FLI.pcp
US-10-425-114-48183

Query Match 9.0%; Score 35; DB 4; Length 437;

```
Best Local Similarity 100.0%; Pred. No. 7.6e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRYWCQKQCRDENGFKCHCMSSHQRQ 53
DB 32 KGLQKLRYWCQKQCRDENGFKCHCMSSHQRQ 66

RESULT 7
US-10-425-114-58982
; Sequence 58982, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 58982
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700238628_FLI.pep
US-10-425-114-58982

Query Match 9.0%; Score 35; DB 4; Length 437;
Best Local Similarity 100.0%; Pred. No. 7.6e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRYWCQKQCRDENGFKCHCMSSHQRQ 53
DB 32 KGLQKLRYWCQKQCRDENGFKCHCMSSHQRQ 66

RESULT 8
US-10-424-599-168326
; Sequence 168326, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 168326
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_123013C.1.pep
US-10-424-599-168326

Query Match 7.7%; Score 30; DB 4; Length 176;
Best Local Similarity 100.0%; Pred. No. 1.7e-19;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRYWCQKQCRDENGFKCHCMSE 48
DB 19 KGLQKLRYWCQKQCRDENGFKCHCMSE 48
```

```
RESULT 9
US-10-106-698-5464
; Sequence 5464, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5464
; LENGTH: 84
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-5464

Query Match 7.4%; Score 29; DB 4; Length 84;
Best Local Similarity 100.0%; Pred. No. 7.6e-19;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 363 FSATIVTGTPLKGRVRVEGIQYEDISKLA 391
DB 56 FSATIVTGTPLKGRVRVEGIQYEDISKLA 84

RESULT 10
US-11-097-143-41040
; Sequence 41040, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41040
; LENGTH: 390
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-41040

Query Match 3.8%; Score 15; DB 6; Length 390;
Best Local Similarity 100.0%; Pred. No. 4.1e-05;
```

```
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 KSKGLQKLRYWQMC 31
    |||||
Db 17 KSKGLQKLRYWQMC 31

RESULT 11
US-10-732-923-11274
; Sequence 11274, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 11274
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Triticum aestivum
US-10-732-923-11274

Query Match 2.3%; Score 9; DB 5; Length 335;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 ELSRENEEE 201
    |||||
Db 318 ELSRENEEE 326

RESULT 12
US-10-739-930-10661
; Sequence 10661, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 10661
; LENGTH: 535
; TYPE: PRT
; ORGANISM: Triticum aestivum
; FEATURE:
; OTHER INFORMATION: Clone ID: TRIAE-23APR03-C568_2.p
US-10-739-930-10661

Query Match 2.3%; Score 9; DB 5; Length 535;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 ELSRENEEE 201
    |||||
Db 518 ELSRENEEE 526

RESULT 13
US-10-732-923-11269
; Sequence 11269, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
```

```
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 11269
; LENGTH: 539
; TYPE: PRT
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(539)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-732-923-11269

Query Match 2.3%; Score 9; DB 5; Length 539;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 ELSRENEEE 201
    |||||
Db 521 ELSRENEEE 529

RESULT 14
US-09-904-987-7
; Sequence 7, Application US/09904987
; Patent No. US20020037908A1
; GENERAL INFORMATION:
; APPLICANT: No. US20020037908Alactyl, Inc.
; TITLE OF INVENTION: Methods and Compositions for Controlling Pathological and Prepathe
; TITLE OF INVENTION: Protein Assembly or Aggregation
; FILE REFERENCE: 42108/26146
; CURRENT APPLICATION NUMBER: US/09/904,987
; CURRENT FILING DATE: 2001-07-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 1543
; TYPE: PRT
; ORGANISM: homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NCBI ENTREZ / XP_003405
; DATABASE ENTRY DATE: 2001-04-16
; RELEVANT RESIDUES: (1)..(1543)
US-09-904-987-7

Query Match 2.3%; Score 9; DB 3; Length 1543;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 TTSKSSSLG 226
    |||||
Db 1219 TTSKSSSLG 1227

RESULT 15
US-10-408-765A-2250
; Sequence 2250, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2250
; LENGTH: 3141
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2250

Query Match 2.3%; Score 9; DB 4; Length 3141;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 218 TTSKSSSLG 226
|||
Db 1219 TTSKSSSLG 1227

Search completed: November 23, 2005, 16:55:48
Job time : 111.717 secs

This Page Blank (usp1c)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:40:00 ; Search time 4.98725 Seconds
(without alignments)
237.628 Million cell updates/sec

Title: US-09-555-529-25
Perfect score: 391
Sequence: 1 MKKSFPLSPKATANKIKG.....GPLKGRVGGIQVEDISKLA 391

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 17545 seqs, 3030971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 17545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA New:
1: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep:*
2: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep:*
3: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep:*
4: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep:*
5: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep:*
6: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep:*
7: /cgn2_6/prodata/2/pubpaa/US11_NEW_PUB.pep:*
8: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	9	2.3	3144	7 US-11-055-035-1	Sequence 1, Appli
2	7	1.8	229	7 US-11-014-629-1	Sequence 1, Appli
3	7	1.8	318	1 US-10-802-796-727	Sequence 727, App
4	7	1.8	334	1 US-10-802-796-728	Sequence 728, App
5	6	1.5	112	1 US-10-793-626-2114	Sequence 2114, Ap
6	6	1.5	149	1 US-10-986-501-231	Sequence 231, App
7	6	1.5	204	1 US-10-793-626-2394	Sequence 2394, Ap
8	6	1.5	215	1 US-10-821-234-1443	Sequence 1443, Ap
9	6	1.5	227	1 US-10-793-626-2642	Sequence 2642, Ap
10	6	1.5	286	7 US-11-137-465-48	Sequence 48, Appl
11	6	1.5	339	1 US-10-821-234-1507	Sequence 1507, Ap
12	6	1.5	346	1 US-10-967-648A-10	Sequence 10, Appl
13	6	1.5	370	1 US-10-821-234-1105	Sequence 1105, Ap
14	6	1.5	389	7 US-11-012-762-72	Sequence 72, Appl
15	6	1.5	391	1 US-10-996-628-2	Sequence 2, Appli
16	6	1.5	398	7 US-11-012-762-74	Sequence 74, Appl
17	6	1.5	441	1 US-10-510-386-162	Sequence 162, App
18	6	1.5	447	1 US-10-821-234-1584	Sequence 1584, Ap
19	6	1.5	458	7 US-11-077-550-114	Sequence 114, App
20	6	1.5	460	1 US-10-858-730-63	Sequence 63, Appl
21	6	1.5	467	1 US-10-510-386-56	Sequence 56, Appl
22	6	1.5	468	7 US-11-076-187-2	Sequence 2, Appli
23	6	1.5	548	7 US-11-077-550-24	Sequence 24, Appl
24	6	1.5	600	1 US-10-131-826A-462	Sequence 462, App
25	6	1.5	633	1 US-10-821-234-1146	Sequence 1146, Ap

26	6	1.5	724	1 US-10-793-626-968	Sequence 968, App
27	6	1.5	749	1 US-10-793-626-138	Sequence 138, App
28	6	1.5	759	1 US-10-858-730-75	Sequence 75, Appl
29	6	1.5	760	1 US-10-858-730-76	Sequence 76, Appl
30	6	1.5	853	1 US-10-821-234-1110	Sequence 1110, Ap
31	6	1.5	858	7 US-11-077-550-22	Sequence 22, Appl
32	6	1.5	860	7 US-11-077-550-175	Sequence 175, App
33	6	1.5	862	7 US-11-077-550-94	Sequence 94, Appl
34	6	1.5	862	7 US-11-077-550-171	Sequence 171, App
35	6	1.5	862	7 US-11-077-550-173	Sequence 173, App
36	6	1.5	864	7 US-11-077-550-102	Sequence 102, App
37	6	1.5	865	7 US-11-077-550-100	Sequence 100, App
38	6	1.5	866	7 US-11-077-550-88	Sequence 88, Appl
39	6	1.5	866	7 US-11-077-550-104	Sequence 104, App
40	6	1.5	867	7 US-11-077-550-80	Sequence 80, Appl
41	6	1.5	867	7 US-11-077-550-96	Sequence 96, Appl
42	6	1.5	867	7 US-11-077-550-98	Sequence 98, Appl
43	6	1.5	870	7 US-11-077-550-92	Sequence 92, Appl
44	6	1.5	871	7 US-11-077-550-84	Sequence 84, Appl
45	6	1.5	871	7 US-11-077-550-86	Sequence 86, Appl

ALIGNMENTS

RESULT 1
US-11-055-035-1
; Sequence 1, Application US/11055035
; Publication No. US20050256072A1
; GENERAL INFORMATION:
; APPLICANT: ARONIN, NEIL
; APPLICANT: ZAMORE, PHILLIP D.
; APPLICANT: BRODERICK, JENNIFER
; TITLE OF INVENTION: DUAL FUNCTIONAL OLIGONUCLEOTIDES FOR USE IN REPRESSING
; TITLE OF INVENTION: MUTANT GENE EXPRESSION
; FILE REFERENCE: UMY-095
; CURRENT APPLICATION NUMBER: US/11/055,035
; CURRENT FILING DATE: 2005-02-09
; PRIOR APPLICATION NUMBER: 60/543,467
; PRIOR FILING DATE: 2004-02-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 1
; LENGTH: 3144
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-055-035-1

Query Match 2.3%; Score 9; DB 7; Length 3144;
Best Local Similarity 100.0%; Pred. No. 0.89;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 TTSSKSSSLG 226
|||
DB 1222 TTSSKSSSLG 1230

RESULT 2
US-11-014-629-1
; Sequence 1, Application US/11014629
; Publication No. US20050244376A1
; GENERAL INFORMATION:
; APPLICANT: Hall, Frederick L.
; APPLICANT: Gordon, Erlinda M.
; TITLE OF INVENTION: TARGETED GENE DELIVERY TO EXTRACELLULAR MATRIX
; FILE REFERENCE: 30863-704.302
; CURRENT APPLICATION NUMBER: US/11/014,629
; CURRENT FILING DATE: 2004-12-15
; PRIOR APPLICATION NUMBER: US 08/837,223
; PRIOR FILING DATE: 1997-04-10
; PRIOR APPLICATION NUMBER: US 09/904,923
; PRIOR FILING DATE: 2001-07-13
; NUMBER OF SEQ ID NOS: 8

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 229
; TYPE: PRT
; ORGANISM: moloney murine leukemia virus
US-11-014-629-1

Query Match
Best Local Similarity 1.8%; Score 7; DB 7; Length 229;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 322 RLKLDQT 328
Db 102 RLKLDQT 108

RESULT 3
US-10-802-796-727
; Sequence 727, Application US/10802796
; Publication No. US20050250104A1
; GENERAL INFORMATION:
; APPLICANT: COLE, STEWART
; APPLICANT: BUCHRIESER-BROSCH, ROLAND
; APPLICANT: GORDON, STEPHEN
; APPLICANT: BILLAULT, ALAIN
; TITLE OF INVENTION: A METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST
; TITLE OF INVENTION: FROM THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED
; TITLE OF INVENTION: DNA LIBRARY. APPLICATION TO THE DETECTION OF
; TITLE OF INVENTION: MYCOBACTERIA.
; FILE REFERENCE: 05394.0011-00000
; CURRENT APPLICATION NUMBER: US/10/802,796
; CURRENT FILING DATE: 2004-03-18
; PRIOR APPLICATION NUMBER: US/09/673,476
; PRIOR FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: PCT/IB99/00740
; PRIOR FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 09/060,756
; PRIOR FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: PatentIn Ver. 2.2
; SEQ ID NO 727
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Mycobacterium sp.
US-10-802-796-727

Query Match
Best Local Similarity 1.8%; Score 7; DB 1; Length 318;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 GAGGSAG 216
Db 202 GAGGSAG 208

RESULT 4
US-10-802-796-728
; Sequence 728, Application US/10802796
; Publication No. US20050250104A1
; GENERAL INFORMATION:
; APPLICANT: COLE, STEWART
; APPLICANT: BUCHRIESER-BROSCH, ROLAND
; APPLICANT: GORDON, STEPHEN
; APPLICANT: BILLAULT, ALAIN
; TITLE OF INVENTION: A METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST
; TITLE OF INVENTION: FROM THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED
; TITLE OF INVENTION: DNA LIBRARY. APPLICATION TO THE DETECTION OF
; TITLE OF INVENTION: MYCOBACTERIA.
; FILE REFERENCE: 05394.0011-00000
; CURRENT APPLICATION NUMBER: US/10/802,796
; CURRENT FILING DATE: 2004-03-18
; PRIOR APPLICATION NUMBER: US/09/673,476
; PRIOR FILING DATE: 2002-03-29
```

```
; PRIOR APPLICATION NUMBER: PCT/IB99/00740
; PRIOR FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 09/060,756
; PRIOR FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: PatentIn Ver. 2.2
; SEQ ID NO 728
; LENGTH: 334
; TYPE: PRT
; ORGANISM: Mycobacterium sp.
US-10-802-796-728

Query Match
Best Local Similarity 1.8%; Score 7; DB 1; Length 334;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 GAGGSAG 216
Db 173 GAGGSAG 179

RESULT 5
US-10-793-626-2114
; Sequence 2114, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2114
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (112)
; OTHER INFORMATION: variable amino acid
US-10-793-626-2114

Query Match
Best Local Similarity 1.5%; Score 6; DB 1; Length 112;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 QRQLLL 56
Db 35 QRQLLL 40

RESULT 6
US-10-986-501-231
; Sequence 231, Application US/10986501
; Publication No. US20050244845A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 90 Human Secreted Proteins
; FILE REFERENCE: P2013P2C1
; CURRENT APPLICATION NUMBER: US/10/986,501
; CURRENT FILING DATE: 2004-11-12
; PRIOR APPLICATION NUMBER: US/10/621,363
; PRIOR FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: 09/969,730
; PRIOR FILING DATE: 2001-10-06
; PRIOR APPLICATION NUMBER: 09/774,639
; PRIOR FILING DATE: 2001-02-01
```



```
; PRIOR APPLICATION NUMBER: 60/238,291
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 09/244,112
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: PCT/US98/16235
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 60/056,371
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,732
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,366
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,364
; PRIOR FILING DATE: 1997-08-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 373
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 231
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-986-501-231

Query Match      1.5%; Score 6; DB 1; Length 149;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      155 LEKKK 160
Db      141 LEKKK 146

RESULT 7
US-10-793-626-2394
; Sequence 2394, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2394
; LENGTH: 204
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; US-10-793-626-2394

Query Match      1.5%; Score 6; DB 1; Length 204;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      51 QRLLL 56
Db      35 QRLLL 40

RESULT 8
US-10-821-234-1443
; Sequence 1443, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom

; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1443
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-821-234-1443

Query Match      1.5%; Score 6; DB 1; Length 215;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      299 KKGVK 304
Db      172 KKGVK 177

RESULT 9
US-10-793-626-2642
; Sequence 2642, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2642
; LENGTH: 227
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; US-10-793-626-2642

Query Match      1.5%; Score 6; DB 1; Length 227;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      353 GTLESI 358
Db      57 GTLESI 62

RESULT 10
US-11-137-465-48
; Sequence 48, Application US/11137465
; Publication No. US20050255558A1
; GENERAL INFORMATION:
; APPLICANT: Agarwal, Pankaj
; APPLICANT: Murdoch, Paul R.
; APPLICANT: Rizvi, Safia, K.
; APPLICANT: Smith, Randall, F.
; APPLICANT: Xiang, Zhaoying
; APPLICANT: Kabnick, Karen
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50018
; CURRENT APPLICATION NUMBER: US/11/137,465
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US/10/239,663
; PRIOR FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/US01/09226
```

```
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/192,158
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: 60/192,668
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/200,166
; PRIOR FILING DATE: 2000-04-27
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 48
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-137-465-48

Query Match      1.5%; Score 6; DB 7; Length 286;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 191 FTLSR 196
Db 172 FTLSR 177

RESULT 11
US-10-821-234-1507
; Sequence 1507, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Presclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt SEQ_genes Version 1.0
; SEQ ID NO 1507
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1507

Query Match      1.5%; Score 6; DB 1; Length 339;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 329 HLETVI 334
Db 94 HLETVI 99

RESULT 12
US-10-967-648A-10
; Sequence 10, Application US/10967648A
; Publication No. US20050245473A1
; GENERAL INFORMATION:
; APPLICANT: Saunders, Nicholas A
; TITLE OF INVENTION: Differentiation- and/or proliferation-modulating agents and uses
; FILE REFERENCE: 12493972
; CURRENT APPLICATION NUMBER: US/10/967,648A
; CURRENT FILING DATE: 2004-10-15
; PRIOR APPLICATION NUMBER: USSN 60/512010
; PRIOR FILING DATE: 2003-10-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 10
; LENGTH: 346
```

```
; TYPE: PRT
; ORGANISM: Human
US-10-967-648A-10

Query Match      1.5%; Score 6; DB 1; Length 346;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 KEVIDR 309
Db 124 KEVIDR 129

RESULT 13
US-10-821-234-1105
; Sequence 1105, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Presclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt SEQ_genes Version 1.0
; SEQ ID NO 1105
; LENGTH: 370
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1105

Query Match      1.5%; Score 6; DB 1; Length 370;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 212 GGSAGA 217
Db 78 GGSAGA 83

RESULT 14
US-11-012-762-72
; Sequence 72, Application US/11012762
; Publication No. US20050244815A1
; GENERAL INFORMATION:
; APPLICANT: Georgia State University Research Foundation, Inc.
; TITLE OF INVENTION: Compositions and Methods for Viral Resistance Genes
; FILE REFERENCE: GSU1.PCT
; CURRENT APPLICATION NUMBER: US/11/012,762
; CURRENT FILING DATE: 2004-12-15
; PRIOR APPLICATION NUMBER: PCT/US03/19300
; PRIOR FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US 60/390,046
; PRIOR FILING DATE: 2002-06-19
; NUMBER OF SEQ ID NOS: 133
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 72
; LENGTH: 389
; TYPE: PRT
; ORGANISM: Bos taurus
US-11-012-762-72

Query Match      1.5%; Score 6; DB 7; Length 389;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 IRROLE 154
Db 149 IRROLE 154
```

Db 101 IRRQLE 106

RESULT 15
US-10-996-628-2
; Sequence 2, Application US/10996628
; Publication No. US20050244415A1
; GENERAL INFORMATION:
; APPLICANT: Hiroshi Ohnishi et al.
; TITLE OF INVENTION: Pharmaceutical composition and method for activating macrophage
; TITLE OF INVENTION: same
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/996,628
; CURRENT FILING DATE: 2004-11-23
; PRIOR APPLICATION NUMBER: JP2004-133599
; PRIOR FILING DATE: 2004-04-28
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-996-628-2

Query Match 1.5%; Score 6; DB 1; Length 391;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 210 GAGGSA 215
|||
Db 43 GAGGSA 48

Search completed: November 23, 2005, 16:52:01
Job time : 5.98725 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:06:33 ; Search time 130.833 Seconds
(without alignments)
1319.820 Million cell updates/sec

Title: US-09-555-529-26
Perfect score: 2037
Sequence: 1 MGKSDFLTPKAIANRIKSG.....GPKLGRVVEGIQVEDISKLA 393

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_21.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*
9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2037	100.0	393	2	AA23620 A human k
2	2037	100.0	393	7	ADJ69659 Human hea
3	1887	92.6	391	2	AA23619 Murine ki
4	1465	71.9	293	2	AA23618 A huamn t
5	1365	67.0	291	2	AA23617 Mouse tru
6	1243	61.0	390	8	ADP22452 Sea-squir
7	1059.5	52.0	390	4	ABE71416 Drosophil
8	978	48.0	411	3	AA23616 Arabidops
9	934	45.9	424	3	AA23615 Maize KIN
10	934	45.9	437	8	ADX96318 Plant ful
11	934	45.9	437	8	ADX78817 Plant ful
12	909.5	44.6	423	3	AB03064 Maize KIN
13	909.5	44.6	423	3	AB03065 Maize KIN
14	909.5	44.6	423	8	ADT58422 Plant pol
15	829	40.7	382	3	AA23614 Arabidops
16	726	35.6	366	3	AA23613 Arabidops
17	615	30.2	340	3	AA23612 Arabidops
18	615	30.2	343	3	AA23611 Arabidops
19	599	29.4	302	3	AA23610 Arabidops
20	370	18.2	84	4	AA23609 Human col
21	176.5	8.7	82	3	AA23608 Arabidops
22	173	8.5	81	3	AA23607 Arabidops
23	165	8.1	76	3	AA23606 Arabidops
24	155	7.6	76	3	AA23605 Arabidops

25	150.5	7.4	278	8	ADX77206
26	145.5	7.1	1114	7	ADJ70313
27	145.5	7.1	1213	4	AA23616 Human hea
28	145.5	7.1	1213	6	ABP97031 Human L-F
29	144.5	7.1	1212	6	ABP97029 Rat L-FIL
30	144	7.1	375	8	AA23614 Arabidops
31	144	7.1	375	8	ADT56172 Plant pol
32	140	6.9	375	3	AA23613 Arabidops
33	140	6.9	450	3	AA23612 Arabidops
34	139	6.8	348	7	AD140582 Human pur
35	138.5	6.8	517	2	AA23611 Arabidops
36	138	6.8	805	8	AA23610 Arabidops
37	137	6.7	323	3	AA23609 Bacterial
38	136	6.7	347	4	AA23608 Arabidops
39	136	6.7	508	7	AD01217 Human nuc
40	136	6.7	508	8	ADN06062 Antipsoi
41	136	6.7	624	5	AA23606 Human RNA
42	136	6.7	624	6	ABU62546 Human PNI
43	135.5	6.7	237	3	AA23605 Arabidops
44	135.5	6.7	246	3	AA23604 Arabidops
45	135.5	6.7	512	7	ADM05395 Human pro

ALIGNMENTS

RESULT 1
AA23620
ID AAY23620 standard; protein; 393 AA.
XX
AC AAY23620;
XX
DT 07-SEP-1999 (first entry)
XX
DE A human kin17 protein.
XX
KW Human; kin17 protein; cell proliferation; fertility;
KW hyperproliferative disease; protein interaction; curved DNA;
KW HIV replication; HIV integration; repair enzyme.
XX
OS Homo sapiens.
XX
PN FR2772046-A1.
XX
PD 11-JUN-1999.
XX
PF 09-DEC-1997; 97FR-00015536.
XX
PR 09-DEC-1997; 97FR-00015536.
(COMS) COMMISSARIAT ENERGIE ATOMIQUE.
Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
WPI; 1999-359999/31.
New DNA coding for human kin17 protein - useful for controlling cell proliferation or fertility.
Claim 19; Page 37-38; 69pp; French.
The present sequence represents a human kin17 protein. The mammalian kin17 protein is useful for preparing a medicament for controlling cell proliferation or for controlling fertility. The medicaments can also be used to treat hyperproliferative diseases. Fragments between amino acids 55 and 235 (preferably between amino acids 129 and 228) of a mammalian kin17 protein are useful for regulating the interaction between proteins and curved DNA. The fragment can be used to block replication of HIV or its integration into the human genome or to target repair enzymes to curved DNA sites. Expression vectors for kin17 can be used for controlling cell proliferation

Sequence 393 AA;


```

PN FR2772046-A1.
XX
XX
PD 11-JUN-1999.
XX
XX 09-DEC-1997; 97FR-00015536.
XX
XX 09-DEC-1997; 97FR-00015536.
XX
XX (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.
XX
XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
XX WPI; 1999-359999/31.
XX
XX New DNA coding for human kin17 protein - useful for controlling cell
XX proliferation or fertility.
XX
XX Claim 19; Page 36-37; 69pp; French.
XX
XX The present sequence represents a murine kin17 protein. The mammalian
XX kin17 protein is useful for preparing a medicament for controlling cell
XX proliferation or for controlling fertility. The medicaments can also be
XX used to treat hyperproliferative diseases. Fragments between amino acids
XX 55 and 235 (preferably between amino acids 129 and 228) of a mammalian
XX kin17 protein are useful for regulating the interaction between proteins
XX and curved DNA. The fragment can be used to block replication of HIV or
XX its integration into the human genome or to target repair enzymes to
XX curved DNA sites. Expression vectors for kin17 can be used for
XX controlling cell proliferation
XX
XX Sequence 391 AA;

Query Match 92.6%; Score 1887; DB 2; Length 391;
Best Local Similarity 92.1%; Pred. No. 2.5e-155;
Matches 362; Conservative 14; Mismatches 15; Indels 2; Gaps 1;

Qy 1 MGKSDFLTPKAIANRIKSKGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQLLASEN 60
Db 1 MGKSDFLTPKAIANRIKSKGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQLLASEN 60

Qy 61 PQQFMDYFSEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREIHNNATQWETLTDFTK 120
Db 61 PQQFMDYFSEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREIHNNATQWETLTDFTK 120

Qy 121 WLGRGLCKVDETPKGWYIQYIDRDPETIRRLQLEKKKKQDLDDEKTAKEFIEEQVRG 180
Db 121 WLGRGLCKVDETPKGWYIQYIDRDPETIRRLQLEKKKKQDLDDEKTAKEFIEEQVRG 180

Qy 181 LEGKEQVPTFTELSENDEKVTNLSKGCSSGATSSKSTLGPALKTIGSSASVK 240
Db 181 LEGKEQVPTFTELSENDEKVTNLSKGCSSGATSSKSTLGPALKTIGSSASVK 240

Qy 241 RKSSQSSQSKKKKKKSSALDEIMEIEEEKRTARTDYWLQPEIIVKIITKKLGEKYHK 300
Db 241 RKSSQSSQSKKKKKKSSALDEIMEIEEEKRTARTDYWLQPEIIVKIITKKLGEKYHK 300

Qy 301 KKAIVKEVIDKYTAVVKMIDSGDKLKDQTHLETVIPAGKRILVLNGVGRNGEGTLESI 360
Db 301 KKAIVKEVIDKYTAVVKMIDSGDKLKDQTHLETVIPAGKRILVLNGVGRNGEGTLESI 360

Qy 361 NEKTFSAITIVETGPKGRRVEGIOYEDISKLA 393
Db 361 NEKTFSAITIVETGPKGRRVEGIOYEDISKLA 393

RESULT 4
AA23618
ID AA23618 standard; protein; 293 AA.
XX
AC AA23618;
XX
XX 07-SEP-1999 (first entry)
XX

```

A huamn truncated kin17 protein.

kin17 protein; cell proliferation; fertility; hyperproliferative disease; protein interaction; curved DNA; HIV replication; HIV integration; repair enzyme.

Homo sapiens.

FR2772046-A1.

11-JUN-1999.

09-DEC-1997; 97FR-00015536.

09-DEC-1997; 97FR-00015536.

(COMS) COMMISSARIAT ENERGIE ATOMIQUE.

Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF; WPI; 1999-359999/31.

New DNA coding for human kin17 protein - useful for controlling cell proliferation or fertility.

Claim 15; Page 34-35; 69pp; French.

The present sequence represents a truncated human kin17 protein with amino acids 129-228 deleted. The mammalian kin17 protein is useful for preparing a medicament for controlling cell proliferation or for controlling fertility. The medicaments can also be used to treat hyperproliferative diseases. Fragments between amino acids 55 and 235 (preferably between amino acids 129 and 228) of a mammalian kin17 protein are useful for regulating the interaction between proteins and curved DNA. The fragment can be used to block replication of HIV or its integration into the human genome or to target repair enzymes to curved DNA sites. Expression vectors for kin17 can be used for controlling cell proliferation

Sequence 293 AA;

Query Match 71.9%; Score 1465; DB 2; Length 293;
Best Local Similarity 74.6%; Pred. No. 9e-119;
Matches 293; Conservative 0; Mismatches 0; Indels 100; Gaps 1;

Qy 1 MGKSDFLTPKAIANRIKSKGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQLLASEN 60
Db 1 MGKSDFLTPKAIANRIKSKGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQLLASEN 60

Qy 61 PQQFMDYFSEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREIHNNATQWETLTDFTK 120
Db 61 PQQFMDYFSEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREIHNNATQWETLTDFTK 120

Qy 121 WLGRGLCKVDETPKGWYIQYIDRDPETIRRLQLEKKKKQDLDDEKTAKEFIEEQVRG 180
Db 121 WLGRGLCKVDETPKGWYIQYIDRDPETIRRLQLEKKKKQDLDDEKTAKEFIEEQVRG 180

Qy 181 LEGKEQVPTFTELSENDEKVTNLSKGCSSGATSSKSTLGPALKTIGSSASVK 240
Db 181 LEGKEQVPTFTELSENDEKVTNLSKGCSSGATSSKSTLGPALKTIGSSASVK 240

Qy 241 RKSSQSSQSKKKKKKSSALDEIMEIEEEKRTARTDYWLQPEIIVKIITKKLGEKYHK 300
Db 241 RKSSQSSQSKKKKKKSSALDEIMEIEEEKRTARTDYWLQPEIIVKIITKKLGEKYHK 300

Qy 301 KKAIVKEVIDKYTAVVKMIDSGDKLKDQTHLETVIPAGKRILVLNGVGRNGEGTLESI 360
Db 301 KKAIVKEVIDKYTAVVKMIDSGDKLKDQTHLETVIPAGKRILVLNGVGRNGEGTLESI 360

Qy 361 NEKTFSAITIVETGPKGRRVEGIOYEDISKLA 393
Db 361 NEKTFSAITIVETGPKGRRVEGIOYEDISKLA 393

RESULT 5
 AAY23617
 ID AAY23617 standard; protein; 291 AA.
 AC AAY23617;
 XX AAY23617;
 DT 07-SEP-1999 (first entry)
 XX Mouse truncated kin17 protein.
 DE
 XX kin17 protein; cell proliferation; fertility; hyperproliferative disease;
 KW protein interaction; curved DNA; HIV replication; HIV integration;
 KW repair enzyme.
 OS
 XX Mus sp.
 PN PR2772046-A1.
 XX
 PD 11-JUN-1999.
 XX
 PF 09-DEC-1997; 97FR-00015536.
 XX
 PR 09-DEC-1997; 97FR-00015536.
 XX
 PA (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
 XX
 PI Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
 XX WPI; 1999-359999/31.
 DR
 XX New DNA coding for human kin17 protein - useful for controlling cell
 PT proliferation or fertility.
 XX
 PS Claim 14; Page 33-34; 69pp; French.
 XX
 CC The present sequence represents a mouse kin17 protein with amino acids
 CC 129-228 deleted. The mammalian kin17 protein is useful for preparing a
 CC medicament for controlling cell proliferation or for controlling
 CC fertility. The medicaments can also be used to treat hyperproliferative
 CC diseases. Fragments between amino acids 55 and 235 (preferably between
 CC amino acids 129 and 228) of a mammalian kin17 protein are useful for
 CC regulating the interaction between proteins and curved DNA. The fragment
 CC can be used to block replication of HIV or its integration into the human
 CC genome or to target repair enzymes to curved DNA sites. Expression
 CC vectors for kin17 can be used for controlling cell proliferation
 XX
 SQ Sequence 291 AA;
 Query Match 67.0%; Score 1365; DB 2; Length 291;
 Best Local Similarity 69.2%; Pred. No. 4.6e-110;
 Matches 272; Conservative 9; Mismatches 10; Indels 102; Gaps 2;
 QY 1 MGKSDFLTPKAIANRIKSKGLQKLRWYCMQCQCRDENGFKCHCMSESHORQLLASEN 60
 DB 1 MGKSDFLSPKAIANRIKSKGLQKLRWYCMQCQCRDENGFKCHCMSESHORQLLASEN 60
 QY 61 PQOFMDYFSEEFNDLELLRRRFGTKRVHNNIVYNEYISHREHHNNATQWETLTDFTK 120
 DB 61 PQOFMDYFSEEFNDLELLRRRFGTKRVHNNIVYNEYISHREHHNNATQWETLTDFTK 120
 QY 121 WLGREGLCKVDYETPKGWYIQYIDRDPETTRRQLELEKKKKQDLDDEKTAFTIEEQVRG 180
 DB 121 WLGREGLC-----
 QY 181 LEGKEQEVPTFTLSRENDEEKTFTNLSKGCSSSGATSSKSTLGPLSALKTIGSSASVK 240
 DB 129 -----ALKLIGSAASGK 140
 QY 241 RKSSQSSSTQSKBKKKKSGALDEIMBIEBKKTARTDYWLQPEIIVKIITTKLGEKYHK 300
 DB 141 RKSSQSSAQ--PAKKKSGALDEIMBIEBKKTARTDQWLQPIVIVKIITTKLGEKYHK 198

QY 301 KKAIVKEVIDKYTAVVVMIDSGDKLKDQTHLETVPAPGKRILVNGGYRGNEGTLSEI 360.
 DB 199 KKGVVKEVIDRYTAVVVMKTDGDRKLDQTHLETVPAPGKRVLVNGGYRGNEGTLSEI 258
 QY 361 NEKTSATIVITETGPKGRVEGIIQYEDISKLA 393
 DB 259 NEKAFSATIVITETGPKGRVEGIIQYEDISKLA 291
 RESULT 6
 ADP22452
 ID ADP22452 standard; protein; 390 AA.
 XX ADP22452;
 AC ADP22452;
 XX
 DT 12-AUG-2004 (first entry)
 XX
 DE Sea-squirt (Ciona intestinalis) zinc finger protein #13.
 XX
 KW sea-squirt; zinc finger protein; gene detection; drug development;
 KW zinc finger protein-associated disease.
 XX
 OS Ciona intestinalis.
 PN JP2004057126-A.
 XX
 PD 26-FEB-2004.
 XX
 PF 31-JUL-2002; 2002JP-00222484.
 XX
 PR 31-JUL-2002; 2002JP-00222484.
 XX
 PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
 XX
 DR WPI; 2004-208711/20.
 DR N-PSDB; ADP22451.
 XX
 PT Novel gene encoding zinc finger protein, useful as probe in gene
 PT detecting instruments and in development of drug for treating zinc finger
 PT protein associated diseases.
 XX
 PS Claim 1; SEQ ID NO 26; 972pp; Japanese.
 CC The invention comprises the amino acid and coding sequences of sea-squirt
 CC (Ciona intestinalis) zinc finger proteins. The DNA and protein sequences
 CC of the invention are useful in a gene detecting instrument. The DNA and
 CC protein sequences of the invention are useful in the development of drugs
 CC for the treatment of zinc finger protein-associated diseases. The present
 CC amino acid sequence represents a sea-squirt zinc finger protein of the
 CC invention.
 XX
 SQ Sequence 390 AA;
 Query Match 61.0%; Score 1243; DB 8; Length 390;
 Best Local Similarity 62.5%; Pred. No. 2.9e-99;
 Matches 247; Conservative 48; Mismatches 86; Indels 14; Gaps 6;
 QY 1 MGKS--DELTTPKAIANRIKSKGLQKLRWYCMQCQCRDENGFKCHCMSESHORQLLAS 58
 DB 1 MGKEKPGFLITAKAIGNRIKSKGLQKLRWYCMQCQCRDENGFKCHCMSESHORQLLIAG 60
 QY 59 ENPOQFMDYFSEEFNDLELLRRRFGTKRVHNNIVYNEYISHREHHNNATQWETLTD 118
 DB 61 ENPGAFHSFFSSFFDFMQLLKTFRGTRVHNNIVYNEYISHREHHNNATKWYLTAF 120
 QY 119 TWLREGGLCKVDYETPKGWYIQYIDRDPETTRRQLELEKKKKQDLDDEKTAFTIEEQVR 178
 DB 121 TWLREGGHCKVDYETPKGWYIQYIDRDLVLARQLELKQKQREDDDERQAKAIEEMVK 180
 QY 179 RGLG-GEQEVPTFTLSRENDEEKTFTNLSKGCSSSGATSSKSTLGPLSALKTIGSSA 237
 DB 181 RQDTAKDSDSGKFTLVLR-NDEEKIAINLQ---MNSEPKAATSSQLAPRV-----SLS 230

Qy	238	SVRKSSQSSTQSEKKEKKKXSALDEIMEIEE-EKGTARTDYLWLOPEIIVKIITKKLGE	299
Db	231	SLVGRKKQPS\$SGFKP\$SR\$KS\$ALDEIIAABEGRKQKQLHTEHLAKNI\$VKIITKRLGD	290
Qy	297	KYHKHKAIVKEVIDKYTAVVKMIDSGDKLKDOTHLETVIPAPGKRILVLNGYRGNEGT	356
Db	291	KYFKKAAVVK\$EQDKFTAI\$VKMIDSGDQVKPDQTHLETVIP\$GKTVLIVNGSYRDCVAT	350
Qy	357	LESINEKTPSATIVTETGPKRRVEGIQYEDISK	391
Db	351	LEKIDTSKFSATLKITEGPLNGRIVEDIKYEDPSK	385
RESULT 7			
ID	ABB71416		
ID	ABB71416	standard; protein; 390 AA.	
AC	ABB71416;		
DT	26-MAR-2002	(first entry)	
DE		Drosophila melanogaster polypeptide SEQ ID NO 41040.	
KW		Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.	
OS		Drosophila melanogaster.	
PN	WO200171042-A2.		
PD	27-SEP-2001.		
PF	23-MAR-2001; 2001WO-US009231.		
PR	23-MAR-2000; 2000US-0191637P.		
PR	11-JUL-2000; 2000US-00614150.		
PA	(PEKE) PE CORP NY.		
PI	Venter JC, Adams M, Li PWD, Myers EW;		
DR	WPI; 2001-656860/75.		
DR	N-PSDB; ABL15519.		
PT	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signaling and cell-cell interactions.		
PT	Disclosure; SEQ ID NO 41040; 21pp + Sequence Listing; English.		
CC	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABLI6176-ABL30511), expressed DNA sequences (ABL01840-ABLI6175) and the encoded proteins (ABBS7737-ABBS7072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences		
CC	Sequence 390 AA;		
Qy	Query Match	52.0%; Score 1059.5; DB 4; Length 390;	
Db	Best Local Similarity	50.7%; Pred. No. 2.8e-83;	
Qy	Matches 207; Conservative	73; Mismatches 91; Indels 37; Gaps 60	
Qy	1	MGKSDFTPKAIANIK\$KGLQKRWYCOMCQCRDENGFKCHCM\$ESHORQLLASEN	60
Db	1	MGR\$EVGTPKVLANKW\$KGLQKRWYCOMCQCRDENGFKCHTM\$ESHORQLL\$PADN	60
Qy	61	PQOFMDYFSEFRND\$FLELLRRRFGTPKRVHNNIVTVNEYISHREIH\$MNATQWETLTDTFK	120

PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145818P.
PR 27-JUL-1999; 99US-0145919P.

PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.

```

PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 28-OCT-1999; 99US-0162142P.

Query Match
Best Local Similarity 48.08; Score 978; DB 3; Length 411;
Matches 201; Conservative 66; Mismatches 123; Indels 24; Gaps 6;

Qy 1 MGKDFLTPKAIANRIKSGIQLRWYQCMQCRDENGFKCHMSHQRQLLASEN 60
Db 1 MGKNDLTPKAIANRIKAKGLQKLWYQCMQCRDENGFKCHMSHQRQMVFGQN 60

Qy 61 PQQMDYFSEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREHHNATOWETLDTFK 120
Db 61 PTRVVDGYSEFEQTFELDLRRSHRFSRIAATVVYNEYINDRHVHNNSTEWATLTFIK 120

Qy 121 WLGRGLCKVDETPKGYIOYIDRDPETIRROLELEKKKODLDEEKTAKFIEEQVRRG 180
Db 121 HLKGTGCKVEETPKGWFYIDRDSQAVKRLKRKIKSDMADDERQERMTARQIERA 180

Qy 181 LE-----GKEQEVPTFELSREND-----KVTFNLSKGACS--SSGATSSK 221
Db 181 AEKLNCGGEGEETSGNDEVDDGDERKXDEDLRLKSGVKGFALGGVQVATGKERGE 240

Qy 222 SSTLPSALKITIGSSASVKRESQSSTQSKKKKXSALDEIMEIEEEK--RTARTDYW 280
Db 241 SSKLLFGDEENDKVERGEKRSKGS--GRSEKERSALDELKMEBEKKKERNMRKDYW 297

Qy 281 LQPIIKVIITKLGEK-YHKKKAIVKEVIDKYTAVVVMIDSGDKLQDTHLETVPAP 339
Db 298 LFEGIIIVKMSKALAEKGYKQKGVKVDYNYVGEIKMLDSKHLVLRVDOKELETVLPOI 357

Qy 340 GKRLVLNGGYRGNEGTLNESINEKTFSATIVETGPKLGRVVEIGIQVEDISKLA 393
Db 358 GGMVKIVNGAYRGNARLLGVDTKEFKCAKQIEKGVVDGRVKSIEYEDICKLA 411

RESULT 9
AAB03063
ID AAB03063 standard; protein; 424 AA.
XX AC AAB03063;
XX DT 27-SEP-2000 (first entry)
XX DE Maize KIN17 orthologue, ZmKINH-1.
XX KW ZmKINH-1; KIN17 orthologue; maize; zinc finger protein; RecA homologue;
XX KW nuclear localisation; nonhomologous recombination;
XX KW illegitimate recombination; double stranded DNA binding; curved DNA;
XX KW homologous gene targeting; transgenic plant.
XX OS Zea mays.
XX PN WO200024900-A1.
XX PD 04-MAY-2000.
XX PF 06-OCT-1999; 99WO-US023280.
XX PR 27-OCT-1998; 98US-0105802P.
XX PA (PION-) PIONEER HI-BRED INT INC.
XX PI Mahajan PB;
XX WPI; 2000-350741/30.
XX DR N-PSDB; AAA52589.
XX Nucleic acids encoding maize KIN17 orthologue proteins useful for

```

preventing illegitimate recombination in cells.

Claim 11; Page 64-66; 84pp; English.

This sequence represents the maize KIN17 orthologue ZmKINH-1. The invention relates to maize KIN17 orthologues (AAB03063-B03065), nucleic acids encoding them (AAA52589-A52591), and expression vectors, transgenic plants and plant seeds comprising nucleotides encoding maize KIN17 orthologues. KIN17 has, until now, been found only in animal (avian, rodent and human) cells, this invention being the first report describing the presence of KIN17 in plants. Murine KIN17 was found to have a significant homology to Escherichia coli RecA protein, and contains a zinc finger motif and a nuclear localisation signal. KIN17 binds double-stranded DNA, preferentially binding to curved DNA, and forms intranuclear foci on overexpression in mammalian cells. It is also induced on exposure to gamma or ultraviolet radiation. These findings indicate that KIN17 plays a role in non-homologous (illegitimate) recombination, which occurs at higher rates among higher eukaryotes, particularly plants. Illegitimate recombination in plants is a major impediment to the generation of transgenic crops such as maize. Maize KIN17 orthologue nucleic acid sequences may be used to generate transgenic plants. The transgenic plants generated can be monocots or dicots and are particularly maize, soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton, rice, barley and millet. In particular, antisense sequences may be used to reduce KIN17 levels in embryogenic callus or embryo cells, thereby reducing the amount of non-homologous recombination and enhancing homologous gene targeting

Sequence 424 AA;

Query Match 45.9%; Score 934; DB 3; Length 424;
 Best Local Similarity 45.3%; Pred. No. 2.7e-72;
 Matches 193; Conservative 75; Mismatches 122; Indels 36; Gaps 7;

```

Qy 1 MGKDFLTPKAIANRIKSGIQLRWYQCMQCRDENGFKCHMSHQRQLLASEN 60
Db 1 MGKHEFLTPTAIANRIKAKGLQKLWYQCMQCRDENGFKCHMSHQRQMVFGMA 60

Qy 61 PQQMDYFSEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREHHNATOWETLDTFK 120
Db 61 PDRVVEGFSEEFLESFLIRRAHRSRVAATVVYNEYIADRHVHNNSTRWATLTFVK 120

Qy 121 WLGRGLCKVDETPKGYIOYIDRDPETIRROLELEKKKODLDEEKTAKFIEEQVRR- 179
Db 121 FLGREGYCKVEDTPKGFMYIDRDSQAVKRLKRKIKSDMADDERQERMTARQIERA 180

Qy 180 -----GLEGEQEVPTFELSRENDDEEKVTNLSKGACSSG--ATSSKSTLGP 227
Db 181 HKSLAKPNGGAAEGEPESEGEYSDDDEQSPEDDSKEADKATGKIALQKAVPGP 240

Qy 228 SALKTIGSSASVKRK-----ESSQSSSTQSKKK-----KKSALDEIMEIEE 269
Db 241 ---KVPFPDPKPKMKFDFBEDSGTRDQEKNELTUKMGKDVKAAEAKRSALDEIMKEEE 297

Qy 270 E-KKRTARTDYWLQPEIIVKIITKLGEK-YHKKKAIVKEVIDKYTAVVVMIDSGDKLKL 327
Db 298 KAKERSNRKDYWLCPGIVKVMKSLAEKGYKQKGVKVKVKKIDKYVGSIEMLSKHLVRV 357

Qy 328 DQTHLETVPAPKRIILVNGGYRGNEGTLNESINEKTFSATIVETGPKLGRVVEIGIQYE 387
Db 358 DQDELETVPQIGGLVRVINGAYRGNARLLGVDTKEFKCAKQVQVEKGLYDGVKVLRAVEYE 417

Qy 388 DISKLA 393
Db 418 DICKIS 423

```

RESULT 10
 ADX96318
 ID ADX96318 standard; protein; 437 AA.

XX AC ADX96318;
 XX

XX New recombinant DNA constructs useful in the field of biochemistry and
PT Genetics, and in particular for producing transgenic plants with improved
PT biological characteristics.

XX Claim 2; SEQ ID NO 6499; 14pp; English.

XX The invention relates a recombinant DNA construct comprising a
CC polynucleotide having any of 5544 nucleotide sequences (CDNAs SEQ ID NO:
CC 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences
CC (SEQ ID NO: 5545-11088). The CDNAs and proteins are from corn, soybean,
CC Arabidopsis, wheat and rape but the specification does not indicate which
CC sequences is derived from which organism. Also included is a method of
CC producing a plant having an improved property, comprising transforming a
CC plant with a recombinant DNA construct comprising a promoter region
CC functional in a plant cell operably joined to a polynucleotide encoding a
CC polypeptide associated with the property, and growing the transformed
CC plant. The property is selected from improving plant cold tolerance, for
CC manipulating growth rate in plant cells by modification of the cell cycle
CC pathway, for improving plant drought tolerance, for providing increased
CC resistance to plant disease, for galactomannan production, for production
CC of plant growth regulators, for improving plant heat tolerance, for
CC improving plant tolerance to herbicides, for increasing the rate of
CC homologous recombination in plants, for lignin production, for improving
CC plant tolerance to extreme osmotic conditions, for improving plant
CC tolerance to pathogens or pests, for yield improvement by modification of
CC photosynthesis, for modifying seed oil yield and/or content, for
CC modifying seed protein yield and/or content, for yield improvement by
CC modification of carbohydrate, nitrogen or phosphorus use and/or uptake
CC and for yield improvement by providing improved plant growth and
CC development under at least one stress condition. The polynucleotide may
CC also encode a plant transcription factor. The methods and compositions of
CC the present invention are useful in the field of biochemistry and
CC genetics, in particular for producing transgenic plants with improved
CC biological characteristics such as increased yield, improved nitrogen
CC flow, increasing plant tolerance to cold or heat, improving plant
CC tolerance to extreme osmotic and drought conditions, and improving plant
CC tolerance to plant pests or pathogens. They can also be used in physical
CC arrays of molecules, plant breeding markers, computer-based storage and
CC analysis systems. The present sequence is one of the 5544 plant protein
CC sequences of the invention. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20040216190.

XX Sequence 423 AA;

Query Match 44.6%; Score 909.5; DB 8; Length 423;
Best Local Similarity 44.6%; Pred. No. 3.6e-70;
Matches 189; Conservative 76; Mismatches 126; Indels 33; Gaps 8;

QY 1 MGKSDFLTPKATANRIKSLQKLWYCOMQKQCRDENGFKCHCMSSHQRLLLASEN 60
DB 1 MGKHEFLTPKATANRIKAKGLQKLWYCOMQKQCRDENGFKCHCMSSHQRLQVFGMA 60
QY 61 PQQFMDYSEFRNDLFLRRRFTKRVHNNIVVNEVISHREHHMNAOWELTDTFK 120
DB 61 PDRVVEGSEIFLESFLSLIRRAHRSVAATVVVNYEADRHVHNNSTRMATLTTEFVK 120
QY 121 WIGRELCKVDTPKGVIOYIDRDPETIRQLELEKKKODDDEEKTAKFIEOVR- 179
DB 121 LLGREGYCKVEPTPKGNWFTYIDRSEQAVKDLRKRIKSDMAEDERQERMIARQIERA 180
QY 180 -----GLEGEQEVPTFTLSRENDEKVTFLNSKGACSSSG--ATSSKSLTG 226
DB 181 HKS LAKPNDGDAEG-EPESGEEYVSGDDGELEEDVSKENATGKIALQATVAG 239
QY 227 PSALTKTIGSSASVKKSSQSSTQSEKK-----KKK-----SALDEIMEIER-E 270
DB 240 PK-VNPFEDKPKVRFDEGLTGCEBKDELAKKKGDKVDKAAARRSALDELAKKEEMA 298
QY 271 KRTARTDYLQPEIIVKTIITKLGK-YHKKAIVKSVIDKYTAVVKMIDSGDKLKLDQ 329

Db 299 KERSNRKDYMLCPGIVVVKVMSKSLAEKGYKQKGVVRKVMYKVGSIEMLESKHVLRVDO 358
QY 330 THLETVIPAGKRILVNGYRGNEGTLESINEKTSATIVISTGPKGRRVREGIOVEDI 389
DB 359 DELETVIPQIGGLVRLINGAYRGSNARLLSVDTKECAKQVQVEKGLYDGKVLRAVKYDI 418
QY 390 SKLA 393
DB 419 CKIS 422

RESULT 15
AAG42577

ID AAG42577 standard; protein; 382 AA.

XX AAG42577;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 53115.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

XX 06-SEP-2000.

PF 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 23-MAR-1999; 99US-0123548P.

PR 25-MAR-1999; 99US-0125788P.

PR 29-MAR-1999; 99US-0126264P.

PR 01-APR-1999; 99US-0126785P.

PR 06-APR-1999; 99US-0128234P.

PR 16-APR-1999; 99US-0128714P.

PR 19-APR-1999; 99US-0130077P.

PR 21-APR-1999; 99US-0130449P.

PR 23-APR-1999; 99US-0130510P.

PR 28-APR-1999; 99US-0131449P.

PR 30-APR-1999; 99US-0132048P.

PR 30-APR-1999; 99US-0132407P.

PR 04-MAY-1999; 99US-0132484P.

PR 05-MAY-1999; 99US-0132485P.

PR 06-MAY-1999; 99US-0132486P.

PR 07-MAY-1999; 99US-0132863P.

PR 11-MAY-1999; 99US-0134256P.

PR 14-MAY-1999; 99US-0134218P.

PR 14-MAY-1999; 99US-0134219P.

PR 14-MAY-1999; 99US-0134221P.

PR 18-MAY-1999; 99US-0134370P.

PR 19-MAY-1999; 99US-0134768P.

PR 20-MAY-1999; 99US-0135124P.

PR 21-MAY-1999; 99US-0135353P.

PR 25-MAY-1999; 99US-0135629P.

PR 27-MAY-1999; 99US-0136021P.

PR 28-MAY-1999; 99US-0136392P.

PR 01-JUN-1999; 99US-0137222P.

PR 03-JUN-1999; 99US-0137528P.

PR 04-JUN-1999; 99US-0137502P.

PR 07-JUN-1999; 99US-0137724P.

PR 08-JUN-1999; 99US-0138094P.

PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 23-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 28-JUN-1999; 99US-0140891P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-01411842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 08-JUL-1999; 99US-0142820P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-01433542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 20-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.

PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 23-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157533P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 40.7%; Score 829; DB 3; Length 382;
Best Local Similarity 45.2%; Pred. No. 3.2e-63;
Matches 174; Conservative 64; Mismatches 123; Indels 24; Gaps 6;
30 MCQKQCRDENGFKCHMSESHORQLLASENPQQFMDYFSEEFNDFLLELRRRFGTKRV 89

Db 1 MCQKQCRDENGFKCHCSESHQRMQVFGQNPTRVVDGYSEEPQTFDLMRSHRFSRI 60
Qy 90 HNNIVYNEYISHREHHMNATOWETLTDFTKWLGRGLCKVDETPKGYIOYIDRDPETI 149
Db 61 AATVYVNEYINDRHHVHMNSTEWATLTFIKHLGKTGCKVEETPKGWFITYIDROSETL 120
Qy 150 RROLELEKKKKQDLDDDEKTAKFTEEQVVRGLE-----GKEQEVPTFTETLSRENDEE--- 201
Db 121 PKERLKNKRVKSDLAEBEEKQREIQRIERAAEKLNGGGEGGETSGNDEVVDGDDDERKK 180
Qy 202 -----KVTFNLSKGACS-SSGATSSKSSSTLGPALKTTIGSSASVKKRESSQSSSTQ 250
Db 181 DEDLRLKSGVKVGFALGGVQVATKGERGESSKLLFGDEENDKVERGEKKRSGDS--- 237
Qy 251 SKEKKKKKSADEIMEIEEEKK-RTARTDYLQPEIIVKIITKYLGBK-YHKKAIVKEV 308
Db 238 GRSEKERRSALDELKMBEEKKKERMNRKDYWLFEGIIIVKMSKALAEKGYKQKGVVYKV 297
Qy 309 IDKYTAVVVMIDSGDKLQDTHLETVI PAPGKRIILVNGGYRGNEGTLESINEKTFSAT 368
Db 298 IDNTVGEIKMLDPSKHVLKVDQKELETVLQITGGMVKIIVNGAYRGSNARLLIGVDTEKFCAK 357
Qy 369 IVIETGPLKRRRVEGIOVEDISKLA 393
Db 358 VOIEKGVYDGRVIXSIEYEDICKLA 382

Search completed: November 23, 2005, 16:28:22
Job time : 132.833 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:19:43 ; Search time 153.39 Seconds
(without alignments)
1807.628 Million cell updates/sec

Title: US-09-555-529-26
Perfect score: 2037
Sequence: 1 MGKSDPLTPKAIANRIKSKG.....GPLKGRVREGIQVEDISKLA 393

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : UniProt.05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2037	100.0	393	060870_HUMAN	Q60870 homo sapien
2	1895	93.0	391	Q8K339_MOUSE	Q8K339 mus musculus
3	1702	83.6	389	Q6GL97_XENTR	Q6GL97 xenopus tro
4	1653	81.1	387	Q68F56_XENLA	Q68F56 xenopus lae
5	1568	77.0	383	Q5RH25_BRARE	Q5RH25 brachydanio
6	1114	54.7	387	Q7PGA1_ANOGA	Q7PGA1 anopheles g
7	1059.5	52.0	390	Q5VPH4_DROME	Q5VPH4 drosophila
8	1057.5	51.9	390	Q8SXR2_DROME	Q8SXR2 drosophila
9	1044	51.3	195	Q9CV58_MOUSE	Q9CV58 mus musculus
10	978	48.0	411	Q9ZVU5_ARATH	Q9ZVU5 arabidopsis
11	952.5	46.8	404	Q9XWF2_CAEEL	Q9XWF2 caenorhabdi
12	936.5	46.0	269	Q4TDV9_TFTNG	Q4TDV9 tetraodon n
13	918	45.1	430	Q75LJ5_ORYSA	Q75LJ5 oryza sativ
14	849.5	41.7	427	Q4S5G2_TETNG	Q4S5G2 tetraodon n
15	845	41.5	402	Q9AY87_ORYSA	Q9AY87 oryza sativ
16	833	40.9	445	Q55D16_DICDI	Q55D16 dictyosteli
17	830	40.7	397	Q60L39_CAEBR	Q60L39 caenorhabdi
18	694.5	34.1	244	Q76926_DROME	Q76926 drosophila
19	659	32.4	378	Q4N9K7_THEPA	Q4N9K7 theileria p
20	655	32.2	441	Q4XDV6_PLACH	Q4XDV6 plasmodium
21	655	32.2	445	Q4Z7D7_PLABE	Q4Z7D7 plasmodium
22	655	32.1	178	Q6X190_DROYA	Q6X190 drosophila
23	652	32.0	412	Q4UG63_THIEA	Q4UG63 theileria a
24	646	31.7	442	Q8IKG2_PLAF7	Q8IKG2 plasmodium
25	640	31.4	457	Q7RD49_PLAYO	Q7RD49 plasmodium
26	593.5	29.1	334	Q4WFG0_ASFFU	Q4WFG0 aspergillus
27	591	29.0	337	Q415W6_GIBBE	Q415W6 gibberella
28	586.5	28.8	304	Q9Y7X9_SCHPO	Q9Y7X9 schizosacch
29	586.5	28.8	335	Q5AQSO_EMENI	Q5AQSO emenillu
30	541	26.6	324	Q55M94_CRYNE	Q55M94 cryptococcu
31	541	26.6	324	Q5K8H7_CRYNE	Q5K8H7 cryptococcu

RESULT 1

O60870_HUMAN					
ID	O60870_HUMAN	PRELIMINARY;	PRT;	393	AA.
AC	O60870;				
DT	01-AUG-1998	(Tremblrel. 07, Created)			
DT	01-JAN-1999	(Tremblrel. 09, Last sequence update)			
DT	13-SRP-2005	(Tremblrel. 31, Last annotation update)			
DE	Kin17 protein (HsKin17 protein) (KIN, antigenic determinant of reca				
DE	protein homolog) (Mouse).				
GN	Name=KIN; Synonyms=Kin17; ORFNames=RP11-264C14.1-001;				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;				
OC	Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	NUCLEOTIDE SEQUENCE.				
RC	TISSUE=Testis;				
RX	MEDLINE=20419742; PubMed=10964102; DOI=10.1093/carcin/21.9.1701;				
RA	Kannouche P., Mauffrey P., Pinon-Lataillade G., Mattei M.G.,				
RA	Sarasin A., Daya-Grosjean L., Angulo J.F.;				
RA	"Molecular cloning and characterization of the human KIN17 cDNA				
RT	encoding a component of the UVC response that is conserved among				
RT	metazoans."				
RL	Carcinogenesis 21:1701-1710(2000).				
RN	[2]				
RP	NUCLEOTIDE SEQUENCE.				
RC	TISSUE=Pancreas;				
RX	MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;				
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,				
RA	Klausner R.D., Colling F.S., Wagner L., Shenmen C.M., Schuler G.D.,				
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,				
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,				
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,				
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,				
RA	Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,				
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,				
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,				
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,				
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,				
RA	Fahy J., Helton E., Kettman M., Madan A.C., Rodriguez S., Sanchez A.,				
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,				
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,				
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,				
RA	Butterfield Y.S., Krzywinski M.I., Skalska U., Smallos D.E.,				
RA	Schneerch A., Schein J.B., Jones S.J.M., Marra M.A.;				
RT	"Generation and initial analysis of more than 15,000 full-length human				
RT	and mouse cDNA sequences."				
RL	proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).				
RN	[3]				
RP	NUCLEOTIDE SEQUENCE.				
RC	TISSUE=Pancreas;				
RA	Strausberg R.;				
RL	Submitted (NOV-2001) to the EMBL/GenBank/DBSJ databases.				

Q525W2 magaportha
Q4PGL1 ustilago ma
Q6C2K1 yarrowia li
Q7E742 neutospora
Q5CFM3 cryptospori
Q5CYD1 cryptospori
Q6BQD0 debaryomyce
Q51F62 entamoeba h
Q759P8 ashbya gos
Q6CXJ7 kluyveromyc
Q5A673 candida alb
Q4XND8 plasmodium
Q6FMS8 candida gla
Q4TDW0 tetraodon n

```
[4]
RP NUCLEOTIDE SEQUENCE.
RA Lovell J.;
RA Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
RL EMBL; AJ005273; CAA06462.1; -; mRNA.
DR EMBL; BC017309; AAH17309.1; -; mRNA.
DR EMBL; AL158044; CAI12959.1; -; Genomic_DNA.
DR Ensemble; ENSG00000151657; Homo sapiens.
DR HGNC; HGNC:6327; KIN.
DR GO; GO:0005634; C:nucleus; TAS.
DR GO; GO:0003677; F:DNA binding; TAS.
DR InterPro; IPR005824; KOW.
DR Pfam; PF00467; KOW; 1.
DR SMART; SM00355; Znf_C2H2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
KW Nuclear protein.
SQ SEQUENCE 393 AA; 45374 MW; 515A89B4C8A4C007 CRC64;

Query Match 100.0%; Score 2037; DB 2; Length 393;
Best Local Similarity 100.0%; Pred. No. 2.2e-104;
Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGKSDFLTPKAIANRIKSKGLQKLRWYCMQCKQCRDENGFKCHCSESHQROL LASEN 60
DB 1 MGKSDFLTPKAIANRIKSKGLQKLRWYCMQCKQCRDENGFKCHCSESHQROL LASEN 60

QY 61 PQQFMDYFSEEPNDFLELLRRRFGTKRVHNNIVVYNEIYSHREHIHMNATOWETLDTFK 120
DB 61 PQQFMDYFSEEPNDFLELLRRRFGTKRVHNNIVVYNEIYSHREHIHMNATOWETLDTFK 120

QY 121 WLGRGLCKVDTPKGYIQYIDRDPETIRROLEKLEKKKQDLDDBEKTAFTIEEQVRG 180
DB 121 WLGRGLCKVDTPKGYIQYIDRDPETIRROLEKLEKKKQDLDDBEKTAFTIEEQVRG 180

QY 181 LEKGEQVPTFTELSRENDEEKTENLSKACSSGATSSKSTGLPSALKTTIGSSASVK 240
DB 181 LEKGEQVPTFTELSRENDEEKTENLSKACSSGATSSKSTGLPSALKTTIGSSASVK 240

QY 241 RKSSQSSSTQSKKKKSSALDEIMEIEEEKKRTARTDYWLQPEIIVKIITKLGEKYHK 300
DB 241 RKSSQSSSTQSKKKKSSALDEIMEIEEEKKRTARTDYWLQPEIIVKIITKLGEKYHK 300

QY 301 KKAIVKEVDKYTAVVKMIDSGDKLKDQTHLETVPAPGKRILVNLGGYRGNEGTLESI 360
DB 301 KKAIVKEVDKYTAVVKMIDSGDKLKDQTHLETVPAPGKRILVNLGGYRGNEGTLESI 360

QY 361 NEKTSATVIEGTGPKGRVEGQYVEDISKLA 393
DB 361 NEKTSATVIEGTGPKGRVEGQYVEDISKLA 393

RESULT 2
O8K339 MOUSE
ID Q8K339_MOUSE PRELIMINARY; PRT; 391 AA.
AC Q8K339;
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Antigenic determinant of rec-A protein.
GN Names:Kin;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loguellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalón D.K., Muzny D.W., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RA Strausberg R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
[3]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=92020193; PubMed=1293796;
RA Angulo J., Rouer E., Mazin A., Mattei M., Tissier A., Horellou P.,
Benarous R., Devoret R.;
RT "Identification and expression of the cDNA of KIN17, a zinc-finger
gene located on mouse chromosome 2, encoding a new DNA binding
protein."
RL Nucleic Acids Res. 19:5117-5123 (1991).
DR EMBL; BC028860; AAH28860.1; -; mRNA.
DR FIR; S18666; S18666.
DR Ensembl; ENSMUSG00000037262; Mus musculus.
DR MGI; MGI:96676; Kin.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR005824; KOW.
DR InterPro; IPR006646; KOW sub.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00467; KOW; 1.
DR SMART; SM00739; KOW; 1.
DR SMART; SM00355; Znf_C2H2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
KW Nuclear protein.
SQ SEQUENCE 391 AA; 44722 MW; 955BC6A4FF4D3B6E CRC64;

Query Match 93.0%; Score 1895; DB 2; Length 391;
Best Local Similarity 92.4%; Pred. No. 1.4e-96;
Matches 363; Conservative 14; Mismatches 14; Indels 2; Gaps 1;

QY 1 MGKSDFLTPKAIANRIKSKGLQKLRWYCMQCKQCRDENGFKCHCSESHQROL LASEN 60
DB 1 MGKSDFLTPKAIANRIKSKGLQKLRWYCMQCKQCRDENGFKCHCSESHQROL LASEN 60

QY 61 PQQFMDYFSEEPNDFLELLRRRFGTKRVHNNIVVYNEIYSHREHIHMNATOWETLDTFK 120
DB 61 PQQFMDYFSEEPNDFLELLRRRFGTKRVHNNIVVYNEIYSHREHIHMNATOWETLDTFK 120

QY 121 WLGRGLCKVDTPKGYIQYIDRDPETIRROLEKLEKKKQDLDDBEKTAFTIEEQVRG 180
DB 121 WLGRGLCKVDTPKGYIQYIDRDPETIRROLEKLEKKKQDLDDBEKTAFTIEEQVRG 180

QY 181 LEKGEQVPTFTELSRENDEEKTENLSKACSSGATSSKSTGLPSALKTTIGSSASVK 240
DB 181 LEKGEQVPTFTELSRENDEEKTENLSKACSSGATSSKSTGLPSALKTTIGSSASVK 240

QY 241 RKSSQSSSTQSKKKKSSALDEIMEIEEEKKRTARTDYWLQPEIIVKIITKLGEKYHK 300
DB 241 RKSSQSSSTQSKKKKSSALDEIMEIEEEKKRTARTDYWLQPEIIVKIITKLGEKYHK 300

QY 301 KKAIVKEVDKYTAVVKMIDSGDKLKDQTHLETVPAPGKRILVNLGGYRGNEGTLESI 360
DB 301 KKAIVKEVDKYTAVVKMIDSGDKLKDQTHLETVPAPGKRILVNLGGYRGNEGTLESI 360
```

Db 299 KGVVKEVIDRYTAVVGVVTDSDGLKLDQTHLETVIPAGKRVFLVNGGYRGNEGTLSEI 358
 Qy 361 NEKTSATVIVETGPKGRRVEGIOVEDISKLA 393
 Db 359 NEKAFSATVIVETGPKGRRVEGIOVEDISKLA 391

RESULT 3
 Q6GL97_XENTR PRELIMINARY; PRT; 389 AA.
 ID Q6GL97; XENTR PRELIMINARY; PRT; 389 AA.
 AC Q6GL97;
 DT 05-JUL-2004 (TReMBLrel. 27, Created)
 DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
 DE MGC69184 protein.
 GN Names=MGC69184;
 OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
 OC Xenopodinae; Xenopus; Silurana.
 OX NCBI_TaxID=8364;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Embryo;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Brownstein M.J., Uddin T.B., Toshitoki S., Carninci P., Prange C.,
 RA Raha S.S., McSwan P.J., McKernan K.J., Casavant T.L., Scheetz T.E.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Pahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Embryo;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Brownstein M.J., Uddin T.B., Toshitoki S., Carninci P., Prange C.,
 RA Raha S.S., McSwan P.J., McKernan K.J., Casavant T.L., Scheetz T.E.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Pahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Embryo;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Brownstein M.J., Uddin T.B., Toshitoki S., Carninci P., Prange C.,
 RA Raha S.S., McSwan P.J., McKernan K.J., Casavant T.L., Scheetz T.E.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Pahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Embryo;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Brownstein M.J., Uddin T.B., Toshitoki S., Carninci P., Prange C.,
 RA Raha S.S., McSwan P.J., McKernan K.J., Casavant T.L., Scheetz T.E.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Pahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Embryo;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Brownstein M.J., Uddin T.B., Toshitoki S., Carninci P., Prange C.,
 RA Raha S.S., McSwan P.J., McKernan K.J., Casavant T.L., Scheetz T.E.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Pahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Embryo;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Brownstein M.J., Uddin T.B., Toshitoki S., Carninci P., Prange C.,
 RA Raha S.S., McSwan P.J., McKernan K.J., Casavant T.L., Scheetz T.E.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Pahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Embryo;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Brownstein M.J., Uddin T.B., Toshitoki S., Carninci P., Prange C.,
 RA Raha S.S., McSwan P.J., McKernan K.J., Casavant T.L., Scheetz T.E.,
 RA Vill


```
Db 121 LGRNGCHVADTDKGFITIDRDPEFLANQKQKQKDDASELAEFIEQVVRGK 180
Qy 182 EGKEQEVPTTETLSRENDEKVTNLSKGACSSGATSKSTLGPAL---KTIGSSASV 239
Db 181 TEBEPCSTGVSSELKRENEEDTIKIELKLG-----SKQQOSTPSAVISKRPFDALDD 231
Qy 240 KRKESSOSTSQSEKKKKKSNALDEIMEIBEKK-RTARTDYWLQPEITIVKIITKLGKEY 298
Db 232 GKXKKTIAATSTNGETKKLSALDELIOEEOQKKNRKYWLAEGLVVKLIRSLGKEY 291
Qy 299 HKKKAIVKEVIDKTYAVVMKIDSGDKLQDTHLETIVPAGKRIILVNGSYRGNECTLE 358
Db 292 YKEGVVVEVEKRYAKIKLETGEKLVQDAHLETIVPAGKQILVNGSYRGCTAVLK 351
Qy 359 SINKETSATIVTETGPKGRVREGIOYEDISKL 392
Db 352 AINTERYSVTETIASGPKGLRLSNVAYEDISKL 385

RESULT 7
Q9VPH4 DROME PRELIMINARY; PRT; 390 AA.
AC Q9VPH4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE CS5649-PA.
GN Name=kin17; ORFName=CG5649;
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
NUCLEOTIDE SEQUENCE:
MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Buesam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fowler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heinen T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA LaSko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moehref J.A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacieb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler P., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
```

```
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA Pacieb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskaas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila
RT melanogaster euchromatic genome sequence.";
RN Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J.W., Svirskaas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomic perspective.";
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Croesby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradscky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Battencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RN Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RG Berkeley Drosophila Genome Project;
RA Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,
RA Hoskins R., Stapleton M., Pacieb J., Park S., Svirskaas R., Smith E.,
RA Yu C., Rubin G.;
RT "Drosophila melanogaster release 4 sequence.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP NUCLEOTIDE SEQUENCE.
RG FlyBase;
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
CC -I- INTERACTION:
CC QSVI16:CG1041; NbExp=1; IntAct=EBI-193999, EBI-175006;
CC QSVPS5:Hsp60B; NbExp=1; IntAct=EBI-193999, EBI-159512;
CC P1187:pAbp; NbExp=1; IntAct=EBI-193999, EBI-103658;
DR EMBL; AE003591; AAF51578.1; -; Genomic_DNA.
DR IntAct; Q9VPH4; -;
DR Ensembl; CG5649; Drosophila melanogaster.
DR FlyBase; FBgn0024887; CG5649.
DR FlyBase; FBgn0024887; kin17.
DR GO; GO:0005634; C:nucleic acid binding; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR007087; Znf_C2H2.
DR SMART; SM00355; Znf_C2H2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
DR Nuclear protein.
SQ SEQUENCE 390 AA; 45355 MW; 5F8231706C2D431F CRC64;
Query Match 52.0%; Score 1059.5; DB 2; Length 390;
Best Local Similarity 50.7%; Pred. No. 1.5e-50;
Matches 207; Conservative 73; Mismatches 91; Indels 37; Gaps 7;
Qy 1 MGKSDFTTPKAIANRIKSKGLQKLRWYQCMQCKCRDENGFKCHMSHQRQLLASEN 60
Db 1 MGRAEVGTGPKYLANRKMKSGLQKLRWYQCMQCKCRDENGFKCHTMSHQRQLLPADN 60
```

```

QY 61 PQOFMDYFSEEFNRDLELLRRRFGTKRVHNNIVYNEYISHREHHHNNATQWETLTDFTK 120
DB 61 PGKFLHSFSEKSGDGYWELLRRRFGTKRTSANKIYQYIAHKEHHHNNATRWLTLSDYVK 120

QY 121 WLGRGLCKVDPTPKGWIQYIDRDPETIRROLELEKKKKQDLDEKTAFTIEEOVR-- 178
DB 121 WLGRGTQVIADETEKGFVYIDRSPEAMERQAKADKEMKEDDERWADFIEQIKNA 180

QY 179 RGLGKEQEB-VPTFTLSRENDEE-KVTFNLSKGACSSGATSSKSTLGPALKTIIGSS 236
DB 179 RGLGKEQEB-VPTFTLSRENDEE-KVTFNLSKGACSSGATSSKSTLGPALKTIIGSS 236

QY 181 KAKDGEEDGQKFTLKRREENEPLKDIRLEK-----KFQDPTVLGKSAL----- 226
DB 181 KAKDGEEDGQKFTLKRREENEPLKDIRLEK-----KFQDPTVLGKSAL----- 226

QY 237 ASVKRKSSQSSTQSKKKKK-----SALDEIMEIEEEKK-RTARTDYWLOPE 284
DB 237 ASVKRKSSQSSTQSKKKKK-----SALDEIMEIEEEKK-RTARTDYWLOPE 284

QY 285 IIVKIITKLGKHYKKKAIKVEVDIKYTAIVKMDSDGKLKLDQTHLETVTPAPCKRIL 344
DB 285 IIVKIITKLGKHYKKKAIKVEVDIKYTAIVKMDSDGKLKLDQTHLETVTPAPCKRIL 344

QY 345 VLVGKFGYRNEGTFLESINEKTSATIVETGPKLGRVVEGIQYEDISKL 392
DB 345 VLVGKFGYRNEGTFLESINEKTSATIVETGPKLGRVVEGIQYEDISKL 392

QY 340 VVNGAYRGSEALLKLDERRYSVSVLEILHGPKLGRIVDNVQYEDISKL 387
DB 340 VVNGAYRGSEALLKLDERRYSVSVLEILHGPKLGRIVDNVQYEDISKL 387

RESULT 8
Q8SXR2 DROME
ID Q8SXR2_DROME PRELIMINARY; PRT; 390 AA.
AC Q8SXR2;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-WAR-2004 (TREMBlrel. 26, Last annotation update)
DE Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310020017 product:antigenic determinant of rec-A protein, full insert sequence. (fragment).
DE Name=Kin;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dreesnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY084184; AAL89922.1; -; mRNA.
DR FlyBase; FBgn0024887; CG5649.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; P:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR007087; Znf_C2H2.
DR SMART; SM00355; Znf_C2H2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
KW Nuclear protein.
SQ SEQUENCE 390 AA; 45286 MW; 39590CD3B61077CC CRC64;

Query Match 51.9%; Score 1057.5; DB 2; Length 390;
Best Local Similarity 50.7%; Pred. No. 2e-50;
Matches 207; Conservative 72; Mismatches 92; Indels 37; Gaps 7;

QY 1 MGKSDFLTPKATNRKSGLOKLRWYCMQCKQDENGFKCHCNESHQROLILASEN 60
DB 1 MGRAEVGTPTLANKWYKSGLOKLRWYCMQCKQDENGFKCHCNESHQROLILASEN 60

QY 61 PQOFMDYFSEEFNRDLELLRRRFGTKRVHNNIVYNEYISHREHHHNNATQWETLTDFTK 120
DB 61 PGKFLHSFSEKSGDGYWELLRRRFGTKRTSANKIYQYIAHKEHHHNNATRWLTLSDYVK 120

```

```

QY 121 WLGRGLCKVDPTPKGWIQYIDRDPETIRROLELEKKKKQDLDEKTAFTIEEOVR-- 178
DB 121 WLGRGTQVIADETEKGFVYIDRSPEAMERQAKADKEMKEDDERWADFIEQIKNA 180

QY 179 RGLGKEQEB-VPTFTLSRENDEE-KVTFNLSKGACSSGATSSKSTLGPALKTIIGSS 236
DB 179 RGLGKEQEB-VPTFTLSRENDEE-KVTFNLSKGACSSGATSSKSTLGPALKTIIGSS 236

QY 181 KAKDGEEDGQKFTLKRREENEPLKDIRLEK-----KFQDPTVLGKSAL----- 226
DB 181 KAKDGEEDGQKFTLKRREENEPLKDIRLEK-----KFQDPTVLGKSAL----- 226

QY 237 ASVKRKSSQSSTQSKKKKK-----SALDEIMEIEEEKK-RTARTDYWLOPE 284
DB 237 ASVKRKSSQSSTQSKKKKK-----SALDEIMEIEEEKK-RTARTDYWLOPE 284

QY 285 IIVKIITKLGKHYKKKAIKVEVDIKYTAIVKMDSDGKLKLDQTHLETVTPAPCKRIL 344
DB 285 IIVKIITKLGKHYKKKAIKVEVDIKYTAIVKMDSDGKLKLDQTHLETVTPAPCKRIL 344

QY 345 VLVGKFGYRNEGTFLESINEKTSATIVETGPKLGRVVEGIQYEDISKL 392
DB 345 VLVGKFGYRNEGTFLESINEKTSATIVETGPKLGRVVEGIQYEDISKL 392

QY 340 VVNGAYRGSEALLKLDERRYSVSVLEILHGPKLGRIVDNVQYEDISKL 387
DB 340 VVNGAYRGSEALLKLDERRYSVSVLEILHGPKLGRIVDNVQYEDISKL 387

RESULT 9
Q9CV58 MOUSE
ID Q9CV58_MOUSE PRELIMINARY; PRT; 195 AA.
AC Q9CV58;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310020017 product:antigenic determinant of rec-A protein, full insert sequence. (fragment).
DE Name=Kin;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning."
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yananaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RA The FANTOM Consortium,

```


RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RL 60,770 full-length cDNAs.";
 RL Nature 420:563-573 (2002).
 [4]

RN NUCLEOTIDE SEQUENCE;
 RC STRAIN=C57BL/6J; TISSUE=Tongue;
 RX MEDLINE=20493374; PubMed=11042159; DOI=10.1101/gr.145100;
 RX Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Konno H., Akizawa K., Nishi K., Kiteunai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsumura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RL "RIKEN integrated sequence analysis (RISA) system-384-format
 RL sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1617-1630 (2000).
 [5]

RN NUCLEOTIDE SEQUENCE;
 RC STRAIN=C57BL/6J; TISSUE=Tongue;
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RX Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,
 RA Konno H., Akizawa K., Nishi K., Kiteunai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsumura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RL "RIKEN integrated sequence analysis (RISA) system-384-format
 RL sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771 (2000).
 [6]

RN NUCLEOTIDE SEQUENCE;
 RC STRAIN=C57BL/6J; TISSUE=Tongue;
 RX Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
 RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
 RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
 RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
 RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai K., Sakai K.,
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shingawa A., Shiraki T.,
 RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
 RA Teijima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
 RA Muramatsu M., Hayashizaki Y.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AK009429; BAB26281.2; -; mRNA.

DR MGI:966776; Kin.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR InterPro; IPR007087; ZnF_C2H2.
 DR SMART; SM00355; ZnF_C2H2; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
 FT NON TER 195 195
 SQ SEQUENCE 195 AA; 23518 MW; 7E112E9CA820BFA6 CRC64;

Query Match 51.3%; Score 1044; DB 2; Length 195;
 Best Local Similarity 98.5%; Pred. No. 5e-50; Indels 0; Gaps 0;
 Matches 192; Conservative 1; Mismatches 2;

QY 1 MGKSDFTLTPKAIANRIKSGKGLKRLWYCMQCKQCRDENGFKCHCMSESHORQLLASSEN 60
 DB 1 MGKSDFTLTPKAIANRIKSGKGLKRLWYCMQCKQCRDENGFKCHCMSESHORQLLASSEN 60
 QY 61 PQQPMDFSEFRNDFLELLRRRFGTKRVHNNIVYNEIYSHREIHNNATOWETLTDTFTK 120
 DB 61 PQQPMDFSEFRNDFLELLRRRFGTKRVHNNIVYNEIYSHREIHNNATOWETLTDTFTK 120
 QY 121 WLGRGLCKVDTPKGVYIQVDRDPETIRRLQLEKKKKQDLDEEKTAKFIEQVRRG 180
 DB 121 WLGRGLCKVDTPKGVYIQVDRDPETIRRLQLEKKKKQDLDEEKTAKFIEQVRRG 180
 QY 181 LEGKEQVPTFTELS 195
 DB 181 LEGKEQVPTFTELS 195

RESULT 10

Q9ZVUS ARATH
 ID Q9ZVUS_ARATH PRELIMINARY; PRT; 411 AA.
 AC Q9ZVUS;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-PBB-2005 (TrEMBLrel. 29, Last annotation update)
 DE TSA14.13 protein (Hypothetical protein At1g55460).
 GN Name:TSA14.13; Synonyms:At1g55460;
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
 RA Altati H., Araujo R., Rowley D., Buehler E., Dunn P.,
 RA Gonzalez A., Kremenetskaia I., Kim C., Lenz C., Li J., Liu S.,
 RA Luros S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,
 RA Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 [2]
 RP NUCLEOTIDE SEQUENCE.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Etgu P., Lee J.M.,
 RA Toriumi M., Yu G., Brooks S., Chao Q., Chen H., Karlin-Neumann G.,
 RA Kim C., Lam B., Miranda M., Nguyen M., Palm C.J., Shinn P.,
 RA Southwick A., Davis R.W., Ecker J.R., Theologis A.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 [3]
 RP NUCLEOTIDE SEQUENCE.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
 RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC005223; AAD10649.1; -; genomic_DNA.
 DR EMBL; AY051011; AAK93688.1; -; mRNA.
 DR EMBL; AF360132; AAK25842.1; -; mRNA.
 DR PIR; H96596; H96596.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR InterPro; IPR005824; KOW.
 DR InterPro; IPR006646; KOW sub.
 DR InterPro; IPR007087; ZnF_C2H2.
 DR Pfam; PF00467; KOW; 1.
 DR SMART; SM00739; KOW; 1.
 DR SMART; SM00355; ZnF_C2H2; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
 KW Hypothetical protein.
 SQ SEQUENCE 411 AA; 47288 MW; 9DA6F8648002065D CRC64;

Query Match 48.0%; Score 978; DB 2; Length 411;
 Best Local Similarity 48.6%; Pred. No. 5e-46;
 Matches 201; Conservative 66; Mismatches 123; Indels 24; Gaps 6;
 QY 1 MGKSDFTLTPKAIANRIKSGKGLKRLWYCMQCKQCRDENGFKCHCMSESHORQLLASSEN 60
 DB 1 MGKSDFTLTPKAIANRIKSGKGLKRLWYCMQCKQCRDENGFKCHCMSESHORQLLASSEN 60
 QY 61 PQQPMDFSEFRNDFLELLRRRFGTKRVHNNIVYNEIYSHREIHNNATOWETLTDTFTK 120
 DB 61 PTVVDGYSSEFEQTFLLMRSHRFRSRIATVYVYNEIYNDRHVHNNSTEWATLTFTK 120
 QY 121 WLGRGLCKVDTPKGVYIQVDRDPETIRRLQLEKKKKQDLDEEKTAKFIEQVRRG 180
 DB 121 WLGRGLCKVDTPKGVYIQVDRDPETIRRLQLEKKKKQDLDEEKTAKFIEQVRRG 180
 DB 121 LGKTKGCKVEETPKGVYITVDRDSETLTKERLKNKRVKSLAESEKQEREIQRIERA 180

```

QY 181 LE-----GKEQVPTFTLSRENDEB-----KYTFNLSKGACS--SSGATSSK 221
DB 181 ABLKNGGGGEGTSGNDEVDGDDDERKKDEDLRLKSGYKVGFGALGGYKVQATCKERGE 240
QY 222 SFTLPSALKTTIGSSAVKRKSSQSSQTSQKSKKKSALDIMEIEBEKK-RTARTDYW 280
DB 241 SKLLFGDEENDKVRGEKKKSGDS--GRSEKERSALDELMKEEKKKERNKRDYW 297
QY 281 LOPEIIVKIITKLGEK-YHKKAIVKEVIDKYTAVVWMIDSGDKLKDQTHLETVPAP 339
DB 298 LPEGIIVKMSKALAEKGYKQGVVKKVIDNVGSEIKMLDSKHVLRVDOKELETVLPQI 357
QY 340 GKRILVNGGYRGNEGTLSEINKEKTSATIVTETGPKGRVVEGQYEDISKLA 393
DB 358 GGMVIVNGAYGRSGNARLLGVDTKFCVAKVIEKGVYDGRVKSIEYEDICKLA 411

RESULT 11
Q9XWF2_CABEL PRELIMINARY; PRT; 404 AA.
AC Q9XWF2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein Y52B11A.9.
GN ORFNames=Y52B11A.9;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916; DOI=10.1126/science.282.5396.2012;
RG The C. elegans sequencing consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; AL032654; CAA21720.1; -; Genomic_DNA.
DR PIR; T27106; T27106.
DR Ensembl; Y52B11A.9; Caenorhabditis elegans.
DR WormBase; WBGene00013128; Y52B11A.9.
DR WormPep; Y52B11A.9; CE20297.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0007275; P:development; IEA.
DR InterPro; IPR002358; Ribosomal_L6_1.
DR InterPro; IPR007087; Znf_C2H2; 1.
DR SMART; SM00355; Znf_C2H2; 1.
DR PROSITE; PS00525; RIBOSOMAL_L6_1; UNKNOWN 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 404 AA; 46401 MW; 9F55A1413525F057 CRC64;

Query Match 46.8%; Score 952.5; DB 2; Length 404;
Best Local Similarity 46.6%; Pred. No. 1.3e-44;
Matches 193; Conservative 80; Mismatches 110; Indels 31; Gaps 8;

QY 1 MCKSDPLTPKATNRITKSGLOKLRWYCOMCKQCDENGFKCHCMSESHORLLLASEN 60
DB 1 MGKHEGSKDKLANRTKSGLOKLRWYCOMCKQCDENGFKCHLTSEAHQRLLLFAEN 60
QY 61 PQOFMDYFEEFRNDPFLLELRFRFGKRVHNNIVVNEYISHREIHNNATQWETLDFPK 120
DB 61 SNSYLRFQSNDEKPNMQLRTSYGKRVANEVYNAFKDGHVHMNSTVWHSLTGFVQ 120
QY 121 WLGREGCLKVDTPKGYIYQYIDRDPETTRRQLEKKKKQDLDDSEKTAFTIEBOVRG 180
DB 121 YLGSSGCKXIDEGDKGYIYIDQ--EALIRKEEDQKQKQKDDERHMQMDGVMQVG 178
QY 181 --LEGKEQVPTFTLSRENDEBKYTFNLSKGACS-----SSGATSSKSTL----- 225

```

```

DB 179 KELAGDDHEYEATELIRDTPDQKIQLDLNLGILDRKLDVKSGVASAKISIFDMPKVKE 238
QY 226 -----GPSALKTISSASVKKKESQSSQTSQKSKKKSALDEIMEIEBEKK-RTARTDY 279
DB 239 DPDEPGPS-----QPSRKSGKRSRSRSPAACKFKSKSALDEIKEMEERKKERNKRDY 292
QY 280 WLOPEIIVKIITKLGEKYYHKKKAIVKEVIDKYTAVVWMIDSGDKLKDQTHLETVPAP 339
DB 293 WREGIIVKVIITKSLGSEYYKAKGVVRKVDVDTAQVK-LDDGTVVKLDDQEHVEVIFSL 351
QY 340 GKRILVNGGYRGNEGTLSEINKEKTSATIVTETGPKGRVVEGQYEDISKLA 393
DB 352 GROMVINGARGQEAATLESIDEXFSLRLKIASPTGREGID-VPYEDASKLA 404

RESULT 12
Q4TDV9_TETNG PRELIMINARY; PRT; 269 AA.
AC Q4TDV9;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Chromosome undetermined SCAF6005, whole genome shotgun sequence.
DE (Fragment).
GN ORFNames=GSTENG0002601001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Jallion O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthonard V., Jubin C., Castellani V., Katinka M., Vacherie B.,
RA Bienmont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
RA Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissenbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CAAE01006005; CAF88923.1; -; Genomic_DNA.
FT NON TER 1
FT NON TER 269
SQ SEQUENCE 269 AA; 31736 MW; E6883839A2CD17AF CRC64;

Query Match 46.0%; Score 936.5; DB 2; Length 269;
Best Local Similarity 69.8%; Pred. No. 6e-44; 32; Indels 9; Gaps 4;
Matches 178; Conservative 36; Mismatches 36;

QY 55 LLASENPOOFMDYFEEFRNDPFLLELRFRFGKRVHNNIVVNEYISHREIHNNATQWET 114
DB 1 LLASENPNRFRMDHFSQEFKDFLELRFRFGKRVHNNIYNEYISDRQIHNNATRWET 60
QY 115 LTDFTKWLGREGCLKVDTPKGYIYQYIDRDPETTRRQLEKKKKQDLDDSEKTAFTIE 174
DB 61 LTDFTKWLGREGCLKVDTPKGYIYQYIDRDPETTRRQLEKKKKQDLDDSEKTAFTIE 120
QY 175 EQVRGLSGKE-QEYPTFTLSRENDEBKYTFNLSKGACSSSGATSSKSTLGPALKTI 233

```

Db	121	EQVRRGDRGTEETPLYTELKENEERKAVFNLGSLCVA--GPSKSPAVGVGALKKE	178
Qy	234	GSASVVRKSSOSSOSQSEKKKKKALDEIMIBEKKR---TARTDWLQPEIIVKII	290
Db	179	AAAASVKRKD--ASSESTKTKRKSGALBIEIEMEKKRKQQOPLRTDYWLQPNIVKVI	235
Qy	291	TKKLGEKYHKKAIV	305
Db	236	TKKLGSRYHKRGV	250
RESULT 13			
Q75LU5	ORYSA		
ID	Q75LU5_ORYSA	PRELIMINARY; PRT; 430 AA.	
AC	Q75LU5;		
DT	05-JUL-2004 (T-EMBLrel. 27, Created)		
DT	05-JUL-2004 (T-EMBLrel. 27, Last sequence update)		
DT	05-JUL-2004 (T-EMBLrel. 27, Last annotation update)		
DE	Hypothetical protein OSJNB0015121.3.		
GN	Names=OSJNB0015121.3;		
OS	Oryza sativa (japonica cultivar-group).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;		
OC	Ehrhartoideae; Oryzeae; Oryza.		
OX	NCBI_TaxID=39947;		
RN	[1]		
RP	NUCLEOTIDE SEQUENCE.		
RA	Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,		
RA	Overton II L.L., Teitlin T., Kim M.M., Bera J.J., Jin S.S.,		
RA	Padrosh D.W., Tallon L.J., Koo H., Ziemann V., Haiao J., Blunt S.,		
RA	Vanaken S.S., Riedmiller S.B., Utterback T.T., Feldblyum T.V.,		
RA	Yang Q.Q., Haas B.J., Suh B.B., Peterson J.J., Quackenbush J.,		
RA	White O., Salzberg S.L., Fraser C.M.;		
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	NUCLEOTIDE SEQUENCE.		
RA	Buell R.;		
RA	Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AC091302; AAR00634.1; -; Genomic_DNA.		
DR	Gramene; Q75LU5; -		
DR	InterPro; IPR005824; K0M.		
DR	InterPro; IPR007087; Znf_C2H2.		
DR	Pfam; PF00467; K0M; 1.		
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.		
KW	Hypothetical protein.		
SQ	SEQUENCE 430 AA; 49067 MW; 6151D748ED2F4690 CRC64;		
Query Match 45.1%; Score 918; DB 2; Length 430;			
Best Local Similarity 45.6%; Pred. No. 1.1e-42;			
Matches 196; Conservative 74; Mismatches 120; Indels 40; Gaps 10;			
Qy	1	MGKSDFLTPKAIANRIKSKGLQKLRYWCQKQKCRDENGFKCHCMSESHORQLLAS	60
Db	1	MKGHEFTLPKAIANRIKAKGLQKLRYWCQKQKCRDENGFKCHCMSESHORQVFGA	60
Qy	61	PQQPMDFSEFRNDFLELLRRRFGTKRVNNIYVYSHREHHNATOWETLTDPTK	120
Db	61	PDRVVEGSBEFLDAFTLLRRARRHSGRIATVYVNEFIADRHVHNNRWATLTFVK	120
Qy	121	WLGREGLCVKDETPKGVYQIVDDPPTIRQLQLEKKKKDDEEKTAKFIEQVRRG	180
Db	121	FLRGEGCHKVDTPKGVFITVYDRDSEQAVKARLKRRIKSDLAEDRQERMIARQERA	180
Qy	181	L-----EGKEQVPTFTLS-RENDEKVTFLNLSKGACSSG--ATSSK	221
Db	181	QOSMGKTNGELGDADSPGSGEGSGSEDEYSDSENHDEGQED-AKENKAAGKIALAQ	239
Qy	222	SSTLGPALKTIGSSAYK--RKSSQSSTQSKEK-----KKK-----KSALDEI	264
Db	240	RAVFGPK-VNPLDDKPKVKGFEEDSVARDKEELAKKKGKDAINAAEARRSALDEL	298
Qy	265	MEIBEE-KGTARTDYWLQPEIIVKIIITKLGEK-YHKKAIVKEVIDKYTAVVMIDSG	322
RESULT 14			
Q4S5G2	TETNG		
ID	Q4S5G2_TETNG	PRELIMINARY; PRT; 427 AA.	
AC	Q4S5G2;		
DT	13-SEP-2005 (T-EMBLrel. 31, Created)		
DT	13-SEP-2005 (T-EMBLrel. 31, Last sequence update)		
DT	13-SEP-2005 (T-EMBLrel. 31, Last annotation update)		
DE	Chromosome 19 SCAF14731, whole genome shotgun sequence.		
GN	ORFNAMES=GSTENG00023750001;		
OS	Tetraodon nigroviridis (Green puffer).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;		
OC	Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;		
OC	Tetraodontidae; Tetraodontidae; Tetraodon.		
OX	NCBI_TaxID=99883;		
RN	[1]		
RP	NUCLEOTIDE SEQUENCE.		
RA	Jailton O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,		
RA	Mauclai E., Bouneau L., Fischer C., Ozouf-Coataz C., Bernot A.,		
RA	Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,		
RA	Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,		
RA	Athouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,		
RA	Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,		
RA	Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,		
RA	Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Boeak S.,		
RA	Kellis M., Volff J.N., Guigo R., Nusbaum C., Kahn D., Robinson-Rechavi M.,		
RA	Lindblad-Toh K., Birren B., Schachter V., Quetier F., Saunin W., Scarpelli C.,		
RA	Laudet V., Lander E.S., Weissbach J., Roest Crollius H.;		
RA	Wincker P., Wesselsbach J., Roest Crollius H.;		
RT	"Genome duplication in the teleost fish Tetraodon nigroviridis reveals		
RT	the early vertebrate proto-karyotype."		
RL	Nature 431:946-957(2004).		
RN	[2]		
RP	NUCLEOTIDE SEQUENCE.		
RG	Genoscope; Whitehead Institute Centre for Genome Research;		
RL	Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.		
CC	-1- CAUTION: The sequence shown here is derived from an		
CC	EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is		
CC	preliminary data.		
DR	EMBL; CAAE01014731; CAG04120.1; -; Genomic_DNA.		
SQ	SEQUENCE 427 AA; 48007 MW; ADFF91299EBCA23B CRC64;		
Query Match 41.7%; Score 849.5; DB 2; Length 427;			
Best Local Similarity 45.3%; Pred. No. 6.3e-39;			
Matches 190; Conservative 50; Mismatches 112; Indels 67; Gaps 9;			
Qy	1	MGKSDFLTPKAIANRIKSKGLQKLRYWCQKQKCRDENGFKCHCMSESHORQLLAS	60
Db	1	MGRADFLSPKAISNRKSKGLQKLRYWCQKQKCRDENGFKCHCMSESHORQLLAS	60
Qy	61	PQQPMDFSEFRNDFLELLRRRFGTKRVNNIYVYSHREHHNATOWETLTDPTK	120
Db	61	PNRFMDHPSQEFKDFLELLRRRFGTKRVNNIYVYSHREHHNATOWETLTDPTK	120
Qy	121	WLGREGLCVKDETPKGVYQIVDDPPTIRQLQLEKKKKDDEEKTAKFIEQVRRG	180
Db	121	WLGREGLCVKDETPKGVYQIVDDPPTIRQLQLEKKKKDDEEKTAKFIEQVRRG	180
Qy	181	LEGKEQBV-----PTFTELSGRENDEEKTFTFLNLSKGACSSGATSSKSTLGPALKT---	232
Db	165	RRGERQVHRGAGPARARRQDGDG-RDAALHGTETGERGVKGQVPRSLFVCSRTLKQSC	223

QY	233	-----ICSSASVKKRSSQSSTQSKKKKKKSSALDEIMEIEEEKKRTARTDYWLOPEI	285
Db	224	CORWRPESSGGVQREERQFGVODGEEVEEVSQGDYRDGGEEEPAAAAAQDGLL-----	279
QY	286	IVKIITTKLGEKYHK--KKAIVKEVIDKYTAVVVMIDSGDKLKL-----	327
Db	280	---AAQHRGQGHQHVRRALQFEERRRHGGEGVFGGEGDRLWRQAEGPDSPOGDHP	336
QY	328	--DQTHLETVIPAPGKRILVLNGGYRGNEGTLSEINEKTFSATIVTETGPKLGRRVVEGI	384
Db	337	CPRQTGADPERSPOGRSSA--GGRNGEE-----LQRHAAAGLWSAEGETRGRGLRGL	387

RESULT 15
Q9AY87 ORYSA
ID Q9AY87 ORYSA PRELIMINARY; PRT; 402 AA.
AC Q9AY87;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Hypothetical protein OSJNBa0004B24.17;
GN Name=OSJNBa0004B24.17;
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Burr P.C., Hsiao J.,
RA Zismann V., Pai G., Bowman C.L., Fujii C.Y., VanAken S.E.,
RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblyum T.V.,
RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Buell R.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC084319; AAG59654.1; -; Genomic_DNA.
DR Gramene; Q9AY87; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR007087; Znf_C2H2.
DR SMART; SM00355; Znf_C2H2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 402 AA; 45962 MW; 1AF4E38ECC60D13E CRC64;

Query Match	41.5%;	Score 845;	DB 2;	Length 402;
Best Local Similarity	43.5%;	Pred. No. 1e-38;		
Matches 187;	Conservative 67;	Mismatches 108;	Indels 68;	Gaps 11;
QY	1	MKGSDFLTPKATANRIKSKGLQKLRWYQCMQKQCRDENGFKCHCMSESHQRLILASEN	60	
Db	1	MKGHEFLTPKATANRIKAKGLQKLRWYQCMQKQCRDENGFKCHCMSESHQRMQVFGQA	60	
QY	61	PQOFWDYFSEEFERNDFLELLRRRFGTKRVNNIVYNEYISHREHIMHNATOWETLTDFTK	120	
Db	61	PDVVGVGSEEFEDAFLLTLRAHRHSRIAATVVVNEFIADRHVHMNSTRWATLTFVK	120	
QY	121	WLREGCLKVDETPKGWYIQYIDRDPETIRROLELEKKKKQDLDDEKTAFTIEEQVRRG	180	
Db	121	FLGRECHCKVEDTPKGMFTYIDRDSQAVKARLKEKRIKSDLAEDQERMIARQIERA	180	
QY	181	L-----EGKEQEVPTFTELS-RENDEEKVTNLSKGACSSG--ATSSK	221	
Db	181	QQSMGKTNGELGDASPDSGESGSEDBYSDSENDHEGQED-AKEANKAAGKIAIALQ	239	
QY	222	SSTLGPALKTKTGSSASVK--RKSSQSSTQSKK-----KKK-----KSALDEI	264	

Search completed: November 23, 2005, 16:33:34
Job time : 155.39 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:23:48 ; Search time 34.0867 Seconds
(without alignments)
953.202 Million cell updates/sec

Title: US-09-555-529-26
Perfect score: 2037
Sequence: 1 MGKSDFLPKATANRIKSG.....GPLKRRVGIQYEDISKLA 393

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5 COMB.pdp.*
2: /cgn2_6/prodata/1/iaa/6 COMB.pdp.*
3: /cgn2_6/prodata/1/iaa/7 COMB.pdp.*
4: /cgn2_6/prodata/1/iaa/8 COMB.pdp.*
5: /cgn2_6/prodata/1/iaa/9 COMB.pdp.*
6: /cgn2_6/prodata/1/iaa/10 COMB.pdp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	379.5	18.6	136	2	US-09-270-767-32008
2	360	17.7	274	2	US-09-248-796A-20041
3	132	6.5	1507	2	US-08-929-329-5
4	127	6.2	2349	2	US-09-538-092-914
5	123	6.0	1055	2	US-09-949-016-9776
6	122.5	6.0	1564	2	US-10-144-198-2
7	122.5	6.0	1564	2	US-10-144-198-4
8	122.5	6.0	2662	2	US-09-595-684B-31
9	122.5	6.0	2663	2	US-09-538-092-1252
10	122	6.0	934	2	US-09-949-016-6779
11	120.5	5.9	793	2	US-09-538-092-1271
12	120.5	5.9	1003	2	US-09-198-452A-17
13	120	5.9	440	2	US-09-762-194-2
14	120	5.9	440	2	US-09-762-194-4
15	120	5.9	586	1	US-08-630-822A-70
16	120	5.9	586	1	US-09-005-069-70
17	120	5.9	586	2	US-09-171-156A-30
18	120	5.9	586	2	US-09-004-730A-30
19	120	5.9	586	2	US-08-981-799A-30
20	120	5.9	1211	2	US-09-134-001C-4820
21	119.5	5.9	1007	2	US-09-155-558-30
22	118.5	5.8	425	2	US-09-286-381B-11
23	118.5	5.8	425	2	US-10-254-995-11
24	118.5	5.8	583	2	US-09-710-279-1358
25	118.5	5.8	671	2	US-09-439-313-380
26	118.5	5.8	671	2	US-09-352-616A-380
27	118.5	5.8	671	2	US-09-289-198-306

28	118.5	5.8	671	2	US-09-636-215-380	Sequence 380, App
29	118.5	5.8	671	2	US-09-685-166A-380	Sequence 380, App
30	118.5	5.8	671	2	US-09-429-755-306	Sequence 306, App
31	118.5	5.8	671	2	US-09-679-426-380	Sequence 380, App
32	118.5	5.8	671	2	US-09-759-143-380	Sequence 380, App
33	118.5	5.8	671	2	US-09-651-236-380	Sequence 380, App
34	118.5	5.8	671	2	US-09-699-295-306	Sequence 306, App
35	118.5	5.8	671	2	US-09-534-825A-306	Sequence 306, App
36	118.5	5.8	671	2	US-09-657-279-380	Sequence 380, App
37	118.5	5.8	671	2	US-10-012-896-380	Sequence 380, App
38	118.5	5.8	1003	2	US-09-438-185A-8	Sequence 8, Appli
39	118.5	5.8	1010	2	US-09-134-001C-5178	Sequence 5178, Ap
40	118.5	5.8	1719	2	US-09-439-313-378	Sequence 378, App
41	118.5	5.8	1719	2	US-09-352-616A-378	Sequence 378, App
42	118.5	5.8	1719	2	US-09-636-215-378	Sequence 378, App
43	118.5	5.8	1719	2	US-09-685-166A-378	Sequence 378, App
44	118.5	5.8	1719	2	US-09-679-426-378	Sequence 378, App
45	118.5	5.8	1719	2	US-09-759-143-378	Sequence 378, App

ALIGNMENTS

RESULT 1
US-09-270-767-32008
; Sequence 32008, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 32008
; LENGTH: 136
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-32008

Query Match 18.6%; Score 379.5; DB 2; Length 136;
Best Local Similarity 54.1%; Pred. No. 1.2e-26;
Matches 72; Conservative 28; Mismatches 32; Indels 1; Gaps 1;

Qy	261	LDIMEIEBKK-RTARTDYWLOPEIIVKLIITKLGKHYKKYKAIYKEVIDKYTAVVKMI	319
Db	1	LDEIIKQESKKERANKDYWLHKGIVWKFISKMGKFFKQKAVLDVIDRYOGKIKFL	60
Qy	320	DSGDKLKDQTHLETVIPAPGRILVNGGYRGNEGTLSEINKEKTSATIVETGPLKGR	379
Db	61	ETGEKUKVDQAHLETVIPALDKPVMVNGAYRSEALLKLDERRYSVSVLEILHGLKGR	120
Qy	380	RVEGIQYEDISKL	392
Db	121	IVDNVQYEDISKL	133

RESULT 2
US-09-248-796A-20041
; Sequence 20041, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13

```
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 20041
; LENGTH: 274
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-20041

Query Match      17.7%; Score 360; DB 2; Length 274;
Best Local Similarity 32.9%; Pred. No. 1.9e-24;
Matches 91; Conservative 52; Mismatches 78; Indels 56; Gaps 9;

QY 1 MCKSDFLPKAIANRIKSGKLGKQKRWYCOMCKQCRDENGPKCHOMCSHQRQL--LLAS 58
Db 36 MAKAEFGTAKYQSKLRAGLQKLPKFCOLCKQCRDNGFKNHLSPLHIKKVSEIHES 95
QY 59 ENPOQFMDYFSEFRNDLFELLRFRFGTKRVHNNIVYNEYISHRBHNNATQWETLTDF 118
Db 96 GDSSKLIETYSTYKFDQKFIKLRIHNGTKFINANKFYQEIYIRERDHIHNNSTRWSLTSP 155
QY 119 TKWLGREGCKV-----DETPGM--YIQYIDRDPETIRROLEKKKKQDLDDEEETA 170
Db 156 IKHLGKNGIVKQTNDESNEBEGFNLEIKLIDRTQTLNAYQIKSGGNSLED-NDEMNDD 214
QY 171 KFIEQVRGLGKEQVPTFTFELSRNDEEKVTFNLSKGACSSSGATSSKSTLGPSPAL 230
Db 215 KLLQKQIKRG-----QEMEKDKDAE-----KSKTLE----- 240
QY 231 KTIGSSASVKRKESQSSSTQSKK---KKKKKSALDE 263
Db 241 QLVPTSPVK-----LTLKKKKTATTCKLVNAFDE 270

RESULT 3
US-08-929-329-5
; Sequence 5, Application US/08929329
; Patent No. 6120770
; GENERAL INFORMATION:
; APPLICANT: Adams, John H
; APPLICANT: Dalton, John P
; APPLICANT: Kappe, Stefan
; TITLE OF INVENTION: Plasmodium Proteins Useful for Preparing
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Barnes & Thornburg
; STREET: 11 S Meridian
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/929,329
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Breen, John P
; REGISTRATION NUMBER: 38,833
; REFERENCE/DOCKET NUMBER: 835910-28685
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 231-7745
; TELEFAX: (317) 231-7433
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1507 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein

; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium yoelii
US-08-929-329-5

Query Match      6.5%; Score 132; DB 2; Length 1507;
Best Local Similarity 19.5%; Pred. No. 0.012;
Matches 67; Conservative 59; Mismatches 131; Indels 86; Gaps 8;

QY 44 HCMSESHQRQLLASENPOQFMDYFSEFRNDLFELLRFRFGTKRVHNNIVYNEYISHRE 103
Db 1046 HNSAPEHFRSLKINSYTPNRRGENFAKE-----SDSTNTDESKWDEVIRKKE 1093
QY 104 HHNNATQWETLTDFTKWLGREGCKVDETPGMWYIQYIDR-----PETIRROLE 154
Db 1094 EAAKNA-----EIIKRPFAQAQAAWAKAEEERKKAEEAEEAEEERKRIE 1138
QY 155 LEKKKKQDLDDEEKTAKFIEQVRGLGKEQVPTFTFELSRNDEEKVTFNLSKGACSS 214
Db 1139 AEKKAEEERKRIEAEKKAEEERKRIEAEKKAEEERKRIEAEKKAEEERKRIEAEKKAEE 1198
QY 215 ----SGATSSKSTLGPSPALKTIGSSASVKRKESQSSSTQSKKKKKSALDEIMEIEBE 270
Db 1199 RKTEAAKKAEEERKKAEEAVK--AEAAKKAEEAAKKAALKESRRGKKTIEAVKKAEE 1256
QY 271 KKR-----TARTDYMLOPEIIVKIITKYLKGLGKRYHKKKAIVK 306
Db 1257 RKRIEAEKKAEEERKRIEAVKQKKKNVENAEKKAEEERKRIEAAKKAEEERIKREAVKK 1316
QY 307 E-----VIDKYTAV-----VKMIDSGDKLKLDQ 329
Db 1317 EEEVVIKKNLSEKKTAIFLKQNSNNYETRNIDNNSPKLDE 1359

RESULT 4
US-09-538-092-914
; Sequence 914, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqformatter Version 0.9
; SEQ ID NO 914
; LENGTH: 2349
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P12270
US-09-538-092-914

Query Match      6.2%; Score 127; DB 2; Length 2349;
Best Local Similarity 21.0%; Pred. No. 0.063;
Matches 69; Conservative 61; Mismatches 114; Indels 84; Gaps 13;

QY 48 ESHQRQLLASENPOQFMDYFSEFR--NDEL-----ELRRRFGTKRVH 90
Db 121 EAEKRDRLTRNERSQEELEYITEDVKRLNEKLNKESNTYKGLQLKLDELQADSVKYRE 180
QY 91 NNIVYNEYISHREHHNNATQWETLTDFTKWLGREGCKVDETPGMWYIQYIDRDPETIR 150
Db 181 KLEQEKELLHSQNTWLN--TELKTKTDELLALGRE-----KGNELDE 221
```

QY 151 RQLELEKKKKQ--DLDE-----EKTAKFIEEOVRGLEGKEQEVPTFTTSLRENDE 200
Db 222 LKCNLENKEEVRLEQOGLKTSNEHLQKQVEDLTLKKAKEQQAASMEKFFHNLNA 281
QY 201 EKVTFLNSKGCSSGATSSKSTLGPALKTIGSSASVKRKSSQSSQSKKKKSA 260
Db 282 HIKLSNLYKSAADSEAKSNELTRAVEELHKL--KEAGEAN-----KAI 324
QY 261 LDIMEIEBKKTARTDWLQPEIIVKI--ITKLG-----EKYHKKAIKVE 307
Db 325 QHLLVEQSKDQ-----MEKEMLEKIGRLEKELENANDLLSATKRKGAILEELAA 377
QY 308 VIDKYTAVVMIDSGDKL-KLDQTHLET 334
Db 378 MSPTAAAVAKIVPGMKLTLYNAYVET 405

RESULT 5
US-09-949-016-9776
; Sequence 9776, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9776
; LENGTH: 1055
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9776

Query Match 6.0%; Score 123; DB 2; Length 1055;
Best Local Similarity 18.6%; Pred. No. 0.045;
Matches 85; Conservative 94; Mismatches 157; Indels 120; Gaps 21;

QY 18 SKGLQKLRYCQCKO-----CRDENGFKCHCMSESHORQLLASENPQQFMDYF 68
Db 516 SKEKEDLOEKCDIWEKLAQTKRVLAANEENS-----KMEQSNLEKLELVNRYKLQQLDQL 571
QY 69 SEE---FRNDFLELLRRRFGTKRVHNNVIVNYISHREHIMN-----AT 110
Db 572 NRDKLSLHND-ISAQQQLQEKREAVNSLOELANVQDHLNLAQDLHTTKHQDVLLSE 630
QY 111 QWETLDTFTWLREGGLCKVDTPKGHYIQYIDRDP-----TIRQLELEKKKKQLDD 165
Db 631 QTRLOKDISEANRFDCQKEETKQOQLQVLQNEIBENKLVQQEMFQRLQKERSE 690
QY 166 EEK--TAKFIEEOVRGLE-----GKEQEVPTFTTSLRENDEERKVTF 205
Db 691 ESKLETSKVLKEQHQHLEKLTDOQSKLDQVLSKVLAAEERVTLQE-----EEWCE 744
QY 206 NLSKGCSSGATSSKSTLGPAA--LKTIGSSASVKRKSS-----QSSTQSKKKKKKS 259
Db 745 SLENTLSQTKQLSERSQQLVEKSGELLALQKQADSMDRDFSLRNQFLTERKKAQKVA 804
QY 260 ALDIMEIE-----EKKKTARTDWLQPEI-IVKIITKLGKHYHKKAIKVEI 309
Db 805 SLKALKIORSQLEKNLLEQKENS-----CIQKEMATIELVAQ---DNHERARRLMKELN 857
QY 310 D---KYTAVVMKI-----DSGDKLKD-----QTHLETV-----IPAGK 341

Db 858 QMQYEYTELKQMANQKOLERRQMBISDAMRTLKSEVKDEIRTSKLKLNOLFELPADLE 917
QY 342 RILVLNGGYRGEGTLESINEXTFSATIVETGPLK 377
Db 918 AILERN---ENLEGELESKE---NLPTMNEGPF 947

RESULT 6
US-10-144-198-2
; Sequence 2, Application US/10144198
; Patent No. 6833247
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies Inc
; TITLE OF INVENTION: Regulated Prostate Cance Genes
; FILE REFERENCE: 9U 105 R1
; CURRENT APPLICATION NUMBER: US/10/144,198
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1564
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-198-2

Query Match 6.0%; Score 122.5; DB 2; Length 1564;
Best Local Similarity 24.2%; Pred. No. 0.089;
Matches 64; Conservative 37; Mismatches 87; Indels 77; Gaps 12;

QY 116 TDFTKWLREGGLCKVDET-----PKG-WYIQYIDRDPETIRROL-ELEKKKKQLDD 165
Db 40 TQCRNWL-KTGNCLYGNTRCFVHGPPSPRGKYSNNYRRSPERTGDLRRMKNKRODVT 98
QY 166 E-----EKTAKFIEEOVRGLEGEQEVPTFTTSLRENDEEKVTFNLSKGACS----- 213
Db 99 EPQKRNTTESSSPVRKESRGRHREKEDIKTKERTPESEENVEWETNRDDSDGDINY 158
QY 214 -----SSGATSSKSTLGPALKTIG 234
Db 159 DYVHELSELMKQKIQRELKMLEQENKXREELIIKKKESPEVVRSKLSP-SFSLRK--- 214
QY 235 SSASVVRKSSQSSQSKKKKKKKSALDEIME-----IEEKKKTARTDWLQPEIIVK 288
Db 215 SSKSPKRKSPKSSSSAKDKRTSAVSSPLLDQORNSKTNQSKKGGPRTP--SPPPPIPE 272
QY 289 IITKLGKHYHKKAIKVEIDKYT 313
Db 273 DIA--LGKKY-KKKYKVDRIEKT 294

RESULT 7
US-10-144-198-4
; Sequence 4, Application US/10144198
; Patent No. 6833247
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies Inc
; TITLE OF INVENTION: Regulated Prostate Cance Genes
; FILE REFERENCE: 9U 105 R1
; CURRENT APPLICATION NUMBER: US/10/144,198
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1564
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-198-4

Query Match 6.0%; Score 122.5; DB 2; Length 1564;
Best Local Similarity 24.2%; Pred. No. 0.089;
Matches 64; Conservative 37; Mismatches 87; Indels 77; Gaps 12;

QY 116 TDFTKWLREGGLCKVDET-----PKG-WYIQYIDRDPETIRROL-ELEKKKKQLDD 165


```
Db 40 TCRAWL-KTGNCLVNTCRFHGSPRGKSSNVRSPERTPTGDLRMRMKNKQDVDT 98
QY 166 E-----EKTAKFIBEQVARGLEGKEQVPTFTLSRENDEKVTFNLSKGACS----- 213
Db 99 EPQKNTBESSPVKRESSGRHREKEDIKITERTPESEENVEWETNRDDSDNGDINY 158
QY 214 -----SSGATSSKSTLSPSALKTTIG 234
Db 159 DYVHELSELMKROKIQRELMLKEQENMEKREBIIIKKEYSPVVRSKLSP-SPSLRK----- 214
QY 235 SSASVKRKESSOSTOSKEKKKKSALDEIMB-----IEBEKRTARTDYWLQPEIIVK 288
Db 215 SKSPKRSKSPSSASKDKORTSAVSSPLDQORNSKTNQSKKGRTP--SPPPPIPE 272
QY 289 IITKLGKEYHKKAIVKEVIDKYT 313
Db 273 DIA--LGKKY-KEKYVKDRIEKT 294

RESULT 8
US-09-595-684B-31
; Sequence 31, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: Vaisberg, Eugeni
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; FILE OF INVENTION: and purifying human kinesins
; FILE REFERENCE: cytop036
; CURRENT APPLICATION NUMBER: US/09/595,684B
; PRIOR FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 2662
; TYPE: PRT
; ORGANISM: Human
US-09-595-684B-31

Query Match 6.0%; Score 122.5; DB 2; Length 2662;
Best Local Similarity 21.2%; Pred. No. 0.2;
Matches 81; Conservative 67; Mismatches 159; Indels 75; Gaps 14;

QY 17 KSKGLQKLRWYQCMQKQCRDENGFKCHCMS-ESHQROLLLASENPQQPMDFSEFRND 75
Db 1104 KHAHKEGELSRTCDRLAEVEEKLKESQQLQEQQLLNQVEEMSEMQKINE----- 1158
QY 76 FLELLRRRFGTKRVHNNIVYNEYISHREHIMNATOWETLTDTKWLGRGLCKVDETPK 135
Db 1159 -IENLKNELKNKELTLEHETERLELAQKLNENYEVKSITKERVKL--KELQSFETER 1215
QY 136 G---WYIQYIDRDPETIRROLE---LEKKKQQLDDE-----EKTAKFIEBQ-VRRGLE 182
Db 1216 DHLRGYIREIATGQTKEELKIAHILKHEQETIDELRRSVSEKTAQIINTQDLEKSHT 1275
QY 183 GKEQVPTFTELS-----RENDEKVTFN-----LSKGACSSSGATSS----- 220
Db 1276 KLQEBIPVLHREQELLPNVKVSQETQTMNELELLTEQSTTKDSTTLARIEMERLRLNEK 1335
QY 221 -----KSSTLGPSALKTIGSSASVKR-----KSSQSTOSQKKEKKKK 258
Db 1336 FOESQEBIKSLTKERDNLTKIKEALEVHQDLKEHIRETLAKIQESQSQKQSLNKKED 1395
QY 259 SALDEIM-EIEBEKRTARTDYWLQPEIIVKIIITKLGKEYHKKAIVKEVIDKYTAVVK 317
Db 1396 NETTKIVSEMEQPKDSAL---LRIEIMGLSKRLQESHDEMSKVAKEKDDLQRLQEV 1452
```

```
QY 318 MIDSGDKLK-----LDQTHLET 334
Db 1453 LQESDQLKENIKEIVAKHLET 1474

RESULT 9
US-09-538-092-1252
; Sequence 1252, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Mansfield, Traci A.
; APPLICANT: Giot, Loic
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqformatter Version 0.9
; SEQ ID NO 1252
; LENGTH: 2663
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)-(0)
; OTHER INFORMATION: Polypeptide Accession Number Q02224
US-09-538-092-1252

Query Match 6.0%; Score 122.5; DB 2; Length 2663;
Best Local Similarity 21.2%; Pred. No. 0.2;
Matches 81; Conservative 67; Mismatches 159; Indels 75; Gaps 14;

QY 17 KSKGLQKLRWYQCMQKQCRDENGFKCHCMS-ESHQROLLLASENPQQPMDFSEFRND 75
Db 1105 KHAHKEGELSRTCDRLAEVEEKLKESQQLQEQQLLNQVEEMSEMQKINE----- 1159
QY 76 FLELLRRRFGTKRVHNNIVYNEYISHREHIMNATOWETLTDTKWLGRGLCKVDETPK 135
Db 1160 -IENLKNELKNKELTLEHETERLELAQKLNENYEVKSITKERVKL--KELQSFETER 1216
QY 136 G---WYIQYIDRDPETIRROLE---LEKKKQQLDDE-----EKTAKFIEBQ-VRRGLE 182
Db 1217 DHLRGYIREIATGQTKEELKIAHILKHEQETIDELRRSVSEKTAQIINTQDLEKSHT 1276
QY 183 GKEQVPTFTELS-----RENDEKVTFN-----LSKGACSSSGATSS----- 220
Db 1277 KLQEBIPVLHREQELLPNVKVSQETQTMNELELLTEQSTTKDSTTLARIEMERLRLNEK 1336
QY 221 -----KSSTLGPSALKTIGSSASVKR-----KSSQSTOSQKKEKKKK 258
Db 1337 FOESQEBIKSLTKERDNLTKIKEALEVHQDLKEHIRETLAKIQESQSQKQSLNKKED 1396
QY 259 SALDEIM-EIEBEKRTARTDYWLQPEIIVKIIITKLGKEYHKKAIVKEVIDKYTAVVK 317
Db 1397 NETTKIVSEMEQPKDSAL---LRIEIMGLSKRLQESHDEMSKVAKEKDDLQRLQEV 1453
QY 318 MIDSGDKLK-----LDQTHLET 334
Db 1454 LQESDQLKENIKEIVAKHLET 1475

RESULT 10
US-09-949-016-6779
; Sequence 6779, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
```


;; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;; FILE REFERENCE: CL001307
;; CURRENT APPLICATION NUMBER: US/09/949,016
;; CURRENT FILING DATE: 2000-04-14
;; PRIOR APPLICATION NUMBER: 60/241,755
;; PRIOR FILING DATE: 2000-10-20
;; PRIOR APPLICATION NUMBER: 60/237,768
;; PRIOR FILING DATE: 2000-10-03
;; PRIOR APPLICATION NUMBER: 60/231,498
;; PRIOR FILING DATE: 2000-09-08
;; NUMBER OF SEQ ID NOS: 207012
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 6779
;; LENGTH: 994
;; TYPE: PRT
;; ORGANISM: Human
US-09-949-016-6779

Query Match 6.0%; Score 122; DB 2; Length 994;
Best Local Similarity 18.6%; Pred. No. 0.051;
Matches 85; Conservative 94; Mismatches 157; Indels 120; Gaps 21;
QY 18 SKGLQKLRYCQMCQKQ-----CRDENGFKCHMSHSHORQLLLASNPQQFMDYF 68
DB 455 SKEKEDLOEKCDIWEKKLAQTKRVLAABENS-----KMEQSNLEKLDLNVKRLQQLDQL 510
QY 69 SEB---FRNDLELLRRRFGTKRVHNNIVVNYSHREHHM-----AT 110
DB 511 NRDKLSLHND-ISAMQOQLQEKREAVNSLQEEELANVQDHLNLAQKQDILLHTTKHQDVLLSE 569
QY 111 QWETLTDFTKWLGREGCLKVDPTPKGWYQYIDRDP-----TIRQLELEKKKKODLDD 165
DB 570 QTRLOKXDISEWANRFDCQKEETKQOQLQVLQNEIEENKLUVOQEMMFQRLQKERESE 629
QY 166 EEK--TAKFTIEQVRGLE-----GREQEYPTFTLSRENDESKVTF 205
DB 630 ESKLETSKVTLKEQOQHQLEKELTDQSKLDQVLSKVLAABERVTLQE-----EERWCE 693
QY 206 NLSKGACSSGATSKSSTLGPSA--LKTIGSSASVKKRESS-----QSSTQSEKKKKKKS 259
DB 684 SLEKTLSTQKRLSREQQVLVEKSGELLALQKEADSNRADFSLLRNQFLTERKKAQKQVA 743
QY 260 ALDEIMEIE-----EKKGTARTDYWLQPEI-IVKIITKLGKHYHKKKAIIVKEVI 309
DB 744 SLKEALKIQRSQLEKNLLEQKQNS---CICKEMATIELVAQ---DNHERARRLMKELN 796
QY 310 D---KYTAVVVKMI-----DSGDKLKLDD-----QTHLETV-----IPAPGK 341
DB 797 QMQUEYTELKQMANQKQLERQWEISDAMRTLKSEVKDEIRTSLNKLNQFLPELPADLE 856
QY 342 RILVNGVYRGNGTLESINEKTFSTATIVIETGPLK 377
DB 857 AILERN--ENLEGELESLEKE---NLPFTWNEGPFE 886

RESULT 11
US-09-538-092-1271
; Sequence 1271, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 1271

;; LENGTH: 793
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number Q05682
US-09-538-092-1271
Query Match 5.9%; Score 120.5; DB 2; Length 793;
Best Local Similarity 17.6%; Pred. No. 0.05;
Matches 65; Conservative 57; Mismatches 125; Indels 123; Gaps 9;
QY 48 ESHQQLLLASNPQQFMDYFSEPRNDFLELLRRRFGTKRVHNNIVVNYSHREHHM 107
DB 105 ERRQKRLQELARQKEFDTTIDASLS-----LPSSRMQNDTAENETTEKEKSES 155
QY 108 NATQWETLTDFTKWLGREGCLKVDPTPKGWYQYIDRDPETIRRLQ-----ELEKKKKQ 161
DB 156 RQERVE-----IET-----ETVTYSYQKNDWRDAENKKE 186
QY 162 DDDDEKTAFTIEQVRGLEQGEVPTFTLSRENDESKVTFNLSKGACSSSGATSSK 221
DB 187 DKEKEEEE---EKPKRGSIGENQVVMVEBKTTSQEBETVMSLKNQOISSEEPKQEE 242
QY 222 SSTLGPSAL-----KITGSSAS----- 238
DB 243 EREGSDEISHHEKMEEDKERAERARLEAEERERIKAEQDKKIADERARIEAEKAA 302
QY 239 -----VKRESSOSTQSEKKKKKKKKALDIMEIEEE 270
DB 303 AQERERRAEERERMRERKRAAEERQRIKEBEKRAAEERQRIKEBEKRAAEERQRIKEE 362
QY 271 KGRTA--RDYWLQPEIIVKIITKLGKHYHKKKAIIVKEVIDKYTAVVVMIDSG--DKLK 326
DB 363 EKRAAEERQRAABEEKAKVBEQKRNKQLEKKKRAMQETKINGEYQKIEGKWNNEKK 422
QY 327 LDQTHLETVI 336
DB 423 AQEDKLQTAIV 432

RESULT 12
US-09-198-452A-17
; Sequence 17, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griflais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention, and treatment of infection
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 17
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: SITE
; LOCATION: 1...1003
; OTHER INFORMATION: Xaaaunknown or other
US-09-198-452A-17

Query Match 5.9%; Score 120.5; DB 2; Length 1003;
Best Local Similarity 20.9%; Pred. No. 0.071;
Matches 81; Conservative 55; Mismatches 155; Indels 97; Gaps 16;
QY 5 DFLTPKAIAIRIKSGLGKQLRWYQMCQKQ--CRDENGFKCH-----CMSHSHORQLLLASE 59
DB 466 DSIFKKIIDN-----FEKLAWFMLSKSICRFTTIIFENHGHGVAKSLHKNVLEK- 518

Qy	60	NPOQFMDYFSEERNDFLELLRRRGTKRVHNNIVY----	NEYISHREHHNNAQTOWETL	115
Dd	519	-----VIYSLOKSRYDIGWSSAKMILHGPNPFFSLEDNKKTIMKEHAEM---	LESLSL	567
Qy	116	TDPFK---WLREGLUCKVDETPKGH-----	YIQ	140
Dd	568	SSYRKVFIALSDENVVDTFSPDKWDLSGIPCRDALSEISRDEQWKKAHLKHQSLEYTQ	627	
Qy	141	YIDR--DPETIRROLELKKKKQDLDBEKTAKFTIEEQVRGLEGEQEVPTFLSRN	198	
Dd	628	ARDRLDOSSKENOKEBAQEYISSWERVKKFIERVQERQIAIQLKPNIEREEST	687	
Qy	199	DEEKVTNFLSGAGCSSCAT-----SSKSSTLGPSALKITIGSSASVKRKESQSSQTQ	250	
Dd	688	TGETVTPTVGTWTASSDLTDILGRIFVSSREDNQOE-----SCVKVLRSHEVMSWE	741	
Qy	251	SKEK--KKKKSALDIMEIBEKKRTARTDYWLQPEIIVKIITKLGEKYHKKAIVREV	308	
Dd	742	VKEVGPKKEFPQDMGSLR-----FFTIEHEELEVLQKDYSKHL--SYFKKVNNKKEV	794	
Qy	309	IDKYTAVVMKDISDGKCLKLDQTHLETVI	336	
Dd	795	--OYAKF-----RLKVLSDLSGLIL	812	

```

RESULT 13
US-09-762-194-2
; Sequence 2, Application US/09762194
; Patent No. 6835539
; GENERAL INFORMATION:
; APPLICANT: Elbaz, Nathalie
; APPLICANT: Nahmias, Clara
; APPLICANT: Strosberg, Arthur Donny
; TITLE OF INVENTION: NUCLEIC SEQUENCES ENCODING AN AT2
; TITLE OF INVENTION: RECEPTOR-INTERACTING PROTEIN (ATIP) AND THEIR APPLICATIONS
; FILE REFERENCE: 33339/208804
; CURRENT APPLICATION NUMBER: US/09/762,194
; CURRENT FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: PCT/FR99/01908
; PRIOR FILING DATE: 1999-08-02
; PRIOR APPLICATION NUMBER: FR 98/09997
; PRIOR FILING DATE: 1998-08-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-762-194-2

```

	Query Match	5.9%;	Score 120;	DB 2;	Length 440;
	Best Local Similarity	20.8%;	Pred. No. 0.023;		
	Matches 66;	Conservative 64;	Mismatches 129;	Indels 58;	Gaps 11
Qy	41	PKCHMSSES----	HQRQILLASENQPPMDYFSESEFRNDFLELLRRRFCTKXVHNNIVYN	96	
Db	74	YKTKCESQSGFILHRLQRLSRGNKFNFEALTVVIIHLLSEREEALKQ-----	HKTL	125	
Qy	97	EYISHREHMHNATOWETLTDTFKWLGREGLCCKVD-ETPKGWYIOVIDRDPETIRROLEL	155		
Db	126	ELVSRGELVVAASSACEKLEK-----	ARADLQYAYQEFVKLNQOHOQOTDRTELEN	175	
Qy	156	EKKKKQDLDBDEKTAFTIEEQVRRGLEGEQEVPTFTELSRENDBEKVTFNLS-----	208		
Db	176	RLKDLYTAECEKLOSIYIEEAKYKTLQEQ-----	FDNLNAAHETTKLEIASHSEKVEL	231	
Qy	209	-KGACSSGATSSKSSTLGPSSALKTIGSSASVKRKESQSSOTSQSEKKKKKKSALDEIMEI	267		
Db	232	LKKTVETSLSEIKKSHMEKKSLDL-----	LNEXQESLEK-QINDLSENDALNRLKS	285	
Qy	268	EEKKRTARTDYWLOPEII-----	VKIITKLGEEKYHK---KKAIVKEVIDKYTAVV	316	
Db	286	EEQKQLSREKANSKPNQWYMLEQLESIAKVLIEIKNEKLHOODMKLMOKRKLVDNNNTALV	345		

```

Qy 317 KNIDSGDKLKLDQTHLE 333
    |||||
Db 346 -----DKLKRQQENE 356

RESULT 14
US-09-762-194-4
; Sequence 4, Application US/09762194
; Patent No. 6835539
; GENERAL INFORMATION:
; APPLICANT: Elbaz, Nathalie
; APPLICANT: Nahmias, Clara
; APPLICANT: Strosberg, Arthur Donny
; TITLE OF INVENTION: NUCLEIC SEQUENCES ENCODING AN AT2
; TITLE OF INVENTION: RECEPTOR-INTERACTING PROTEIN (ATIP) AND THEIR APPLICATIONS
; FILE REFERENCE: 33339/208804
; CURRENT APPLICATION NUMBER: US/09/762,194
; CURRENT FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: PCT/FR99/01908
; PRIOR FILING DATE: 1999-08-02
; PRIOR APPLICATION NUMBER: FR 98/09997
; PRIOR FILING DATE: 1998-08-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-762-194-4

```

RESULT 15
US-08-630-822A-70
; Sequence 70, Application US/08630822A
; Patent No. 5840695
; GENERAL INFORMATION:
; APPLICANT: FRANK, GLENN R.
; APPLICANT: HUNTER, SHIRLEY WU
; APPLICANT: WALLENFELS, LINDA
; TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS
; TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P. C.
;

Search completed: November 23, 2005, 16:35:43
Job time : 36.0867 secs

This Page Blank (used)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:24:04 ; Search time 114.291 Seconds
(without alignments)
1436.746 Million cell updates/sec

Title: US-09-555-529-26
Perfect score: 2037
Sequence: 1 MGKSDFLTPKAIANRIKSGK.....GPLKGRVBSIQYEDISKLA 393

Scoring table: BLOSUM62
Gapop 10,0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.Main:
1: /cgn2_6/ptodata/1/pubpa/US07_PUBCOMB.pcp:*
2: /cgn2_6/ptodata/1/pubpa/US08_PUBCOMB.pcp:*
3: /cgn2_6/ptodata/1/pubpa/US09_PUBCOMB.pcp:*
4: /cgn2_6/ptodata/1/pubpa/US10A_PUBCOMB.pcp:*
5: /cgn2_6/ptodata/1/pubpa/US10B_PUBCOMB.pcp:*
6: /cgn2_6/ptodata/1/pubpa/US11_PUBCOMB.pcp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2037	100.0	393	4	US-10-408-765A-1465
2	1059.5	52.0	390	6	US-11-097-143-41040
3	934	45.9	437	4	US-10-425-114-48183
4	934	45.9	437	4	US-10-425-114-48183
5	929	45.6	424	4	US-10-425-115-341788
6	909.5	44.6	423	5	US-10-739-930-8499
7	617	30.3	176	4	US-10-424-599-168326
8	535	26.3	138	4	US-10-437-963-191916
9	453	22.2	120	4	US-10-767-701-57151
10	370	18.2	84	4	US-10-106-698-5464
11	326	16.0	162	4	US-10-424-599-196166
12	239	11.7	123	4	US-10-425-115-300949
13	222.5	10.9	107	4	US-10-425-115-230934
14	209	10.3	109	4	US-10-424-599-196168
15	150.5	7.4	278	4	US-10-425-114-46572
16	147.5	7.2	430	4	US-10-425-115-284883
17	145.5	7.1	1114	5	US-10-408-765A-2119
18	144.5	7.1	1213	5	US-10-788-793-6
19	144.5	7.1	1212	5	US-10-788-793-2
20	144	7.1	375	5	US-10-739-930-6249
21	138	6.8	805	4	US-10-369-433-152
22	136	6.7	624	4	US-10-105-959-4
23	136	6.7	624	4	US-10-258-662-20
24	135.5	6.7	512	4	US-10-108-260A-4080
25	135.5	6.7	921	5	US-10-732-923-3305
26	133	6.5	1359	5	US-10-732-923-8707
27	133	6.5	1359	5	US-10-732-923-8708

ALIGNMENTS

RESULT 1

US-10-408-765A-1465
; Sequence 1465, Application US/10408765A
; Publication No. US20040101874A1

GENERAL INFORMATION:

; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1465
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-408-765A-1465

Query Match 100.0%; Score 2037; DB 4; Length 393;

Best Local Similarity 100.0%; Pred. No. 2.8e-143;

Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MGKSDFLTPKAIANRIKSGKGLQKLRWYCMQCKQCRDENGPKCHCMESHORQLLAS	60
Db	1	MGKSDFLTPKAIANRIKSGKGLQKLRWYCMQCKQCRDENGPKCHCMESHORQLLAS	60
Qy	61	PQPMDFSEFRNDFLELLRRRGTGRVNNIYVYSHREHNNATOWELTDFTK	120
Db	61	PQPMDFSEFRNDFLELLRRRGTGRVNNIYVYSHREHNNATOWELTDFTK	120
Qy	121	WLGREGKCKVDTEPKGYIQYIDRDPETIRRLQLEKKKKQDLDEEKTAKFIEQVRRG	180
Db	121	WLGREGKCKVDTEPKGYIQYIDRDPETIRRLQLEKKKKQDLDEEKTAKFIEQVRRG	180
Qy	181	LEGKEQVPTTELSENDEBEKVTFNLSKGACSSSGATSSKSTLGPSSALKTIGSSASVK	240
Db	181	LEGKEQVPTTELSENDEBEKVTFNLSKGACSSSGATSSKSTLGPSSALKTIGSSASVK	240
Qy	241	RKSSQSSQTSQSEKSKKKSALDEIMEIEEKKRTARTDYWLQPEIIVKIITKLGKGYHK	300
Db	241	RKSSQSSQTSQSEKSKKKSALDEIMEIEEKKRTARTDYWLQPEIIVKIITKLGKGYHK	300
Qy	301	KKAIIVKIDKYTAVVXKM1DSGDKLKDQTHLETVIPAPGKRILVLNGGYRGNEGTLSS1	360

Db 301 KKAIVKEVDKTTAVVWMDSDGKLKDDTHLETVIPAPCKRLLVNGYRGNGTLES1 360
Qy 361 NEKTSATIVETGPKLGRVREGIOYEDISKLA 393
Db 361 NEKTSATIVETGPKLGRVREGIOYEDISKLA 393

RESULT 2

US-11-097-143-41040
; Sequence 41040, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41040
; LENGTH: 390
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-41040

Query Match 52.0%; Score 1059.5; DB 6; Length 390;
Best Local Similarity 50.7%; Pred. No. 1.7e-70;
Matches 207; Conservative 73; Mismatches 91; Indels 37; Gaps 7;
Qy 1 MGKSDFLTPKAIANRIKSKGLQKLRYWCQKQCRDENGFKCHCMSESHORQLLASN 60
Db 1 MGRAEVTGPKYLANMKSKGLQKLRYWCQKQCRDENGFKCHCMSESHORQLLASN 60
Qy 61 PQOFMDYFSEEPFNDLELLRRRFGTKRVHNNIVVNEYSHREHIHMNATOWETLTDFTK 120
Db 61 PGKFLHSFKSFSDGQWELLRRRFGTKRSANKIYQEIYAHKEHIHMNATRLWLTSDYVK 120
Qy 121 WLGRGLCKVDTPKGYIYQIDRDPETIRRLQLELEKKKKQDLDDEEKTAKFIEQVR-- 178
Db 121 WLGRGLCKVDTPKGYIYQIDRDPETIRRLQLELEKKKKQDLDDEEKTAKFIEQVR-- 178
Qy 121 WLGRGLCKVDTPKGYIYQIDRDPETIRRLQLELEKKKKQDLDDEEKTAKFIEQVR-- 178
Db 121 WLGRGLCKVDTPKGYIYQIDRDPETIRRLQLELEKKKKQDLDDEEKTAKFIEQVR-- 178
Qy 179 RGLGKQEB--VPTFTLSRENDEE--KVTNLSKGACSSGATSSSTLGPALKTIGSS 236
Db 181 KAKDGEDEGEQKFTBLKRENEPLKDIRLEK-----KFPDPTVLGKSA----- 226
Qy 237 ASVKRESQSSTQSKKKKK-----SALDEIMEIEKKK--RTARDYWLQPE 284
Db 227 -----AKRPAPAEAEKVFPPKPSVAGDSQTSVLDEIIOKEESKKEKRNKDYWLHG 279
Qy 285 IIVKIITKGLKHYHKKAIKVEVIDKYTAVVWMDSDGKLKDDTHLETVIPAPKRLIL 344
Db 280 IIVKFTISKMGKFFKQKAVLVLDIVDYOQKIKFLETGKGLKYDQAHLETVIPALDKPVM 339

Qy 345 VLVNGYRGNGTLESINKEKTSATIVETGPKLGRVREGIOYEDISKL 392
Db 340 VVNGAYRGSEALLRKLDERRYSVVEILHGLPKGRIVDNVQVEDISKL 387
RESULT 3
US-10-425-114-48183
; Sequence 48183, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 48183
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3591-011-E3_FLI1.pep
US-10-425-114-48183

Query Match 45.9%; Score 934; DB 4; Length 437;
Best Local Similarity 45.3%; Pred. No. 4.4e-61;
Matches 193; Conservative 75; Mismatches 122; Indels 36; Gaps 7;
Qy 1 MGKSDFLTPKAIANRIKSKGLQKLRYWCQKQCRDENGFKCHCMSESHORQLLASN 60
Db 14 MGKHEFLTPKAIANRIKAKGLQKLRYWCQKQCRDENGFKCHCMSESHORQMVFGMA 73
Qy 61 PQOFMDYFSEEPFNDLELLRRRFGTKRVHNNIVVNEYSHREHIHMNATOWETLTDFTK 120
Db 74 PDVVVEGFSSEFLSFLSIRRAHRSVAATVVVYVNEIADRRHHVHMNSTRAWLTTEFYK 133
Qy 121 WLGRGLCKVDTPKGYIYQIDRDPETIRRLQLELEKKKKQDLDDEEKTAKFIEQVR-- 179
Db 134 FLGREGYCKVEDTPKGFWMFYIDRSEQAVKRLKRRKIKSDMADDERQERMIARQIERA 193
Qy 180 -----GLEGKEQEVPTFTLSRENDEEKVTFNLSKGACSSG--ATSSKSTLGP 227
Db 194 HKSLAKPNNGGAAEGEPESGSEYSGSDDDDEQEPEDDSKEADKATGKIAIALQKAVPQP 253
Qy 228 SALKTIGSSASVKRK-----ESSQSSTQSKKKK-----KKSALDEIMEIEE 269
Db 254 ---KVPFPDDPKPMKFGFDEEDSGTRDQKNELTKKMGKDVKAAAEAKRSALDELMEKEE 310
Qy 270 E-KKTARTDYWLQPEIIVKIITKGLGK--YHKKAIKVEVIDKYTAVVWMDSDGKLK 327
Db 311 KAKERSNRKDYWLQPEIIVKVMKSLAEKGYIVVKGWKKVYIDKYVGETEMESKHVLRV 370
Qy 328 DQTHLETVIPAPKRIKLVNGYRGNGTLESINKEKTSATIVETGPKLGRVREGIOY 387
Db 371 DQDELETVIPQIGGLVIRVNGAYRGSNARLLSVDTKEKCAKQVQEKGLYDGKVLRAVEYE 430
Qy 388 DISKLA 393
Db 431 DICKIS 436

RESULT 4

US-10-425-114-58982
; Sequence 58982, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong

APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E
APPLICANT: Tabaska, Jack E
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53313)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 58982
LENGTH: 437
TYPE: PRT
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: 700238628_FLI.ppep
US-10-425-114-58982

Query Match 45.9%; Score 934; DB 4; Length 437;
Best Local Similarity 45.3%; Pred. No. 4.4e-61;
Matches 193; Conservative 75; Mismatches 122; Indels 36; Gaps 7;
QY 1 MGKSDFLTPKAIANRIKSKGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQLLASEN 60
DB 14 MGKHEFLTPKAIANRIKAKGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQVFGMA 73
QY 61 PQQFMDYFSEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREHHNNATQWETLTDFTK 120
DB 74 PDRVGEFSEFLESFLSLIRRAHRSRVAATVVYNEYIADRHVHNNSTRWATLTFEVK 133
QY 121 WLREGCLKVDTPKGYIYIDRDPTIRROLELEKKKKQDLDDEKTAKEEQQVRR- 179
DB 134 FLGREGYCKVEDTPKGFMTYIDRDSEQAVKRLKRIKSDMADDERQERMIARQIERA 193
QY 180 -----GLEKGEQVPTFTELSRNDEEKVTNLSKGACSSG--ATSSKSSTLGP 227
DB 194 HKSLAKPNGGGAEGEPESGEYSDDDEQEPEDDSKEADKATGKIAIAIQKAVPGP 253
QY 228 SALKTTIGSSASVKK-----SSOSSTQSKEKK-----KKSALDEIMEIEE 269
DB 254 ---KVPFDDPKPMKGFDEEDSGTRDQEKNELTKMGKDVKAEEKRSALDELMEKEE 310
QY 270 E-KKRTARTDYWLOPEIIVKIITKLGK-YHKKAIVKEVIDKYTAVVMIDSGDKLKL 327
DB 311 KAKERSNRKDYWLCPGIIVKVMKSLAEKGYKQGVVKKVIDKYVGEIEMLESKHVLRV 370
QY 328 DQTHLETVIPAPKRIILVNGYRGNEGTLSEINKEFTSATIVETGPKGRRVEGIOYE 387
DB 371 DQDELETVIPQIGGLVRVINGAYRGSNARLLSVDTFCAKQVQKGLYDGKVLRAVEYE 430
QY 388 DISKLA 393
DB 431 DICKIS 436

RESULT 5
US-10-425-115-341788
Sequence 341788, Application US/10425115
Publication No. US20040214272A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
TITLE OF INVENTION: Plants
FILE REFERENCE: 38-21(53322)B
CURRENT APPLICATION NUMBER: US/10/425,115
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 369326
SEQ ID NO 341788
LENGTH: 424

TYPE: PRT
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: MRT4577_74879C.1.ppep
US-10-425-115-341788
Query Match 45.6%; Score 929; DB 4; Length 424;
Best Local Similarity 45.3%; Pred. No. 1e-60;
Matches 193; Conservative 73; Mismatches 124; Indels 36; Gaps 7;
QY 1 MGKSDFLTPKAIANRIKSKGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQLLASEN 60
DB 1 MGKHEFLTPKAIANRIKAKGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQVFGMA 60
QY 61 PQQFMDYFSEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREHHNNATQWETLTDFTK 120
DB 61 PDRVGEFSEFLESFLSLIRRAHRSRVAATVVYNEYIADRHVHNNSTRWATLTFEVK 120
QY 121 WLREGCLKVDTPKGYIYIDRDPTIRROLELEKKKKQDLDDEKTAKEEQQVRR- 179
DB 121 LLGREGYCKVEDTPKGFMTYIDRDSEQAVKRLKRIKSDMADDERQERMIARQIERA 180
QY 180 -----GLEKGEQVPTFTELSRNDEEKVTNLSKGACSSG--ATSSKSSTLGP 227
DB 181 HKSLAKPNGGGAEGEPESGEYSDDDEQEPEDDSKEADKATGKIAIAIQKAVPGP 240
QY 228 SALKTTIGSSASVKK-----SSOSSTQSKEKK-----KKSALDEIMEIEE 269
DB 241 ---KVPFDDPKPMKGFDEEDSGTRDQEKNELTKMGKDVKAEEKRSALDELMEKEE 297
QY 270 E-KKRTARTDYWLOPEIIVKIITKLGK-YHKKAIVKEVIDKYTAVVMIDSGDKLKL 327
DB 298 KAKERSNRKDYWLCPGIIVKVMKSLAEKGYKQGVVKKVIDKYVGEIEMLESKHVLRV 357
QY 328 DQTHLETVIPAPKRIILVNGYRGNEGTLSEINKEFTSATIVETGPKGRRVEGIOYE 387
DB 358 DQDELETVIPQIGGLVRVINGAYRGSNARLLSVDTFCAKQVQKGLYDGKVLRAVEYE 417
QY 388 DISKLA 393
DB 418 DICKIS 423
RESULT 6
US-10-739-930-8499
Sequence 8499, Application US/10739930
Publication No. US20040216190A1
GENERAL INFORMATION:
APPLICANT: Kovalic, David K.
TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
FILE REFERENCE: 38-21(53377)B
CURRENT APPLICATION NUMBER: US/10/739,930
CURRENT FILING DATE: 2003-12-18
NUMBER OF SEQ ID NOS: 11088
SEQ ID NO 8499
LENGTH: 423
TYPE: PRT
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: ZEAMA-23APR03-C7583_2.p
US-10-739-930-8499
Query Match 44.6%; Score 909.5; DB 5; Length 423;
Best Local Similarity 44.6%; Pred. No. 2.8e-59;
Matches 189; Conservative 76; Mismatches 126; Indels 33; Gaps 8;
QY 1 MGKSDFLTPKAIANRIKSKGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQLLASEN 60
DB 1 MGKHEFLTPKAIANRIKAKGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQVFGMA 60
QY 61 PQQFMDYFSEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREHHNNATQWETLTDFTK 120

```
Db 61 PDRVVEGSEFLESLIRRAHRSRVAATVVYNEYIADRRHHVHMNSTRWATLTFVK 120
Qy 121 WLGRGLCKVDTPKGWYIYIDRDPETIRRQLELEKXKQDLDDDEKTAFTIEQVRR- 179
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 121 LLRGEGYCKVEDTPKGWFTYIDRSEQAVKDLKRIKSDMAEDERQERMIARQIERA 180
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 180 -----GLEGEQEVPTFTELSRENDEKVTFNLSKGCSSG--ATSSKSSTLG 226
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 181 HKSLAKPNDGDAEG-EPESGGEYSGDDGDELEDVSKANKATGKIAIALQTAVRG 239
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 227 PSALKTIGSSASVKRKESSQTSQKEKK----KKK-----SALDIMEIEE-E 270
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 240 PK-VNPFDEKPEVKRFPDEGLGTGEEDKDELAKKGKDVKAADARRSALDELMKEEWA 298
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 271 KRTARTDWLOPELIIVKIITKLGEK-YHKKAIVKEVIDKYTAVVKMIDSGDKLKLQ 329
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 299 KERSNRKDWLCPGIVKVMKSLAEKGYKQGVVRKVMQKYGVEIMLESKHVLRVDQ 358
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 330 THLETVIPAGKRILVNGGYRGNEGTLSEINEKTFSATIVTETGLKRRVEGTYEDI 389
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 359 DELETVIPQIGVLRLNGAYRGSNARLLSVDTEKFCAKVQVEKGLYDGKVLRAVKYDI 418
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 390 SKLA 393
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 419 CKIS 422
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 7
US-10-424-599-168326
; Sequence 168326, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 168326
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_123013C.1.pep
US-10-424-599-168326

Query Match 30.3%; Score 617; DB 4; Length 176;
Best Local Similarity 64.9%; Pred. No. 5.5e-38;
Matches 109; Conservative 25; Mismatches 34; Indels 0; Gaps 0;

Qy 1 MGKSFLLTPKAIANRIKSGLOKLRYWCQKQCRDENGFKCHCMSESHQRLILASEN 60
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1 MGKNEFLTPKAIANRIKAKGLQKLRYWCQKQCRDENGFKCHCMSEGHQRMQIFGN 60
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 61 PQOFMDYFSEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREHIHMNATOWETLTDFTK 120
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 61 PRIVEGYSEEFESTPLEMKHSRFSRVAATVVYNEYINDRRHHIMNSTQWATLTFVK 120
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 121 WLGRGLCKVDTPKGWYIYIDRDPETIRRQLELEKXKQDLDDDEK 168
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 121 YLGRGTCKCKVEETPKGWFYIDRSDSETLTKERMKNKRIKADWDVEK 168
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 8
US-10-437-963-191916
; Sequence 191916, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
```

```
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 191916
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_88190C.1.pep
US-10-437-963-191916

Query Match 26.3%; Score 535; DB 4; Length 138;
Best Local Similarity 67.4%; Pred. No. 5.2e-32;
Matches 93; Conservative 21; Mismatches 24; Indels 0; Gaps 0;

Qy 1 MGKSFLLTPKAIANRIKSGLOKLRYWCQKQCRDENGFKCHCMSESHQRLILASEN 60
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1 MGKHEFLTPKAIANRIKAKGLQKLRYWCQKQCRDENGFKCHCMSESHQRMQVFGQA 60
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 61 PQOFMDYFSEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREHIHMNATOWETLTDFTK 120
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 61 PDRVVEGSEFLEFLDALTLRLRAHRSRIAATVVTNEFIADRRHHVHMNSTRWATLTFVK 120
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 121 WLGRGLCKVDTPKGWY 138
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 121 FLGRSGHCKVEDTPKGWF 138
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 9
US-10-767-701-57151
; Sequence 57151, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 57151
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 30963818.pep
US-10-767-701-57151

Query Match 22.2%; Score 453; DB 4; Length 120;
Best Local Similarity 67.5%; Pred. No. 5.5e-26;
Matches 81; Conservative 14; Mismatches 25; Indels 0; Gaps 0;

Qy 1 MGKSFLLTPKAIANRIKSGLOKLRYWCQKQCRDENGFKCHCMSESHQRLILASEN 60
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1 MGKHEFLTPKAIANRIKAKGLQKLRYWCQKQCRDENGFKCHCMSESHQRMQVFGWA 60
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 61 PQOFMDYFSEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREHIHMNATOWETLTDFTK 120
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 61 PDRVVEGSEFLESFLSLIRRAHRSRVAATVVYNEYIADRRHHVHMNSTRWATLTFVK 120
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
```



```
RESULT 10
US-10-106-698-5464
; Sequence 5464, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5464
; LENGTH: 84
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-5464

Query Match      18.2%; Score 370; DB 4; Length 84;
Best Local Similarity 98.6%; Pred. No. 5.3e-20;
Matches 73; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 320 DSGDKLKLDTHTLETVPAPGRKILVNGYRGNEGTLSEINKEFTSATIVETGPKLGR 379
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 11 ESGDKLKLDTHTLETVPAPGRKILVNGYRGNEGTLSEINKEFTSATIVETGPKLGR 70

QY 380 RVEGIQVEDISKLA 393
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 71 RVEGIQVEDISKLA 84

RESULT 11
US-10-424-599-196166
; Sequence 196166, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 196166
; LENGTH: 162
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_19163C.1.pep
US-10-424-599-196166

Query Match      16.0%; Score 326; DB 4; Length 162;
Best Local Similarity 43.7%; Pred. No. 2.3e-16;
Matches 69; Conservative 32; Mismatches 45; Indels 12; Gaps 3;

QY 238 SVKEKSSQSQTQSEKKKKSALDEIMEIEKK-RTARDYWLQPEIIVKIITKKLGE 296
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 15 NKKKESGG-----KSALDEMMREKKKKNKDYWLHGEIVVKNVKVLAE 64

QY 297 K-YHKKAIVKEVDIKYAVVKMIDSGDKLKDTHLETVPAPGRKILVNGYRGNEG 355
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 65 KGYKQGVVRKIDKYVGEIEMLESKHLRVDDQAELETVPQVGGRVKVNNGAYRGSA 124

QY 356 TLESINEKFTSATIVETGPKLGRRVEGIQVEDISKLA 393
```

```
Db 125 KLLGVDTDNFCAKVQIEKGAYDGRVLKSVYEYEDICKVA 162

RESULT 12
US-10-425-115-300949
; Sequence 300949, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 300949
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(123)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_37535C.1.pep
US-10-425-115-300949

Query Match      11.7%; Score 239; DB 4; Length 123;
Best Local Similarity 46.5%; Pred. No. 4.9e-10;
Matches 46; Conservative 14; Mismatches 39; Indels 0; Gaps 0;

QY 10 KAIANRIKSKGLQKLRYWCOMCQCRDENGFKCHCHSESHQRLLIASENPQOFMYFS 69
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 25 KAIANRIKRRGLQKLQXCHMCHRRDQKELKCHCMRETHQRNMHVLMGMSDTRVXGFS 84

QY 70 EFRNDFLELLRRRFGTKRVHNNIVYNEYISHREHIMN 108
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 85 KEFLXSFLSFIRHADRHSRLTATXVYNYKYIVDRYHVMN 123

RESULT 13
US-10-425-115-230934
; Sequence 230934, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 230934
; LENGTH: 107
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_142205C.1.pep
US-10-425-115-230934

Query Match      10.9%; Score 222.5; DB 4; Length 107;
Best Local Similarity 39.6%; Pred. No. 7e-09;
Matches 42; Conservative 29; Mismatches 34; Indels 1; Gaps 1;

QY 289 IITKKLAEK-YHKKAIVKEVDIKYAVVKMIDSGDKLKDTHLETVPAPGRKILVNL 347
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```

Db      1  VMKSLAEKYYKQGVRRKMKDYGEIEMLESKHVLRVDQDELETVPQIGLLRLN 60

Qy      348  GYRGNGEGLTLESINEKTFSAITVIETGLKRRVEGIQYEDISKLA 393
           |||:  |:::  |:::  |:::  |:::  |:::  |:::  |:::
Db      61  GAYRGSNARLLSVDTEKFCAKQVEKGLYDGKVLRAVEYEDICKIS 106

RESULT 14
US-10-424-599-196168
; Sequence 196168, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 196168
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(109)
; OTHER INFORMATION: unsure at all xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_19165C.1.pep
US-10-424-599-196168

```

[illegible]

```

RESULT 15
US-10-425-114-46572
; Sequence 46572, Application US/10425114
; Publication No, US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 46572
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700442544_FLI.pep
US-10-425-114-46572

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:28:34 ; Search time 5.01276 Seconds
(without alignments)
237.628 Million cell updates/sec

Title: US-09-555-529-26
Perfect score: 2037
Sequence: 1 MKSDFLTPKAIANKRIKSG.....GPLKGRVEGIQVEDISKLA 393

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 17545 seqs, 3030971 residues

Total number of hits satisfying chosen parameters: 17545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA New.*
1: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB_PEP.*
2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB_PEP.*
3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB_PEP.*
4: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB_PEP.*
5: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB_PEP.*
6: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB_PEP.*
7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB_PEP.*
8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB_PEP.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	118.5	5.8	583	1	US-10-793-626-1358
2	111.5	5.5	885	1	US-10-793-626-1660
3	106.5	5.2	2897	1	US-10-499-715-2
4	106	5.2	1189	7	US-11-074-176-134
5	104.5	5.1	1448	1	US-10-485-517-212
6	101.5	5.0	5024	1	US-10-793-626-2964
7	101	5.0	674	1	US-10-507-275-9
8	100	4.9	1299	1	US-10-821-234-1145
9	99	4.9	745	1	US-10-793-626-1742
10	98.5	4.8	752	1	US-10-793-626-348
11	95.5	4.7	676	7	US-11-135-855-28
12	95.5	4.7	717	7	US-11-135-855-29
13	95	4.7	422	1	US-10-821-234-1313
14	95	4.7	1155	1	US-10-793-626-1780
15	94.5	4.6	763	1	US-10-821-234-1619
16	94	4.6	1432	1	US-10-510-386-218
17	93	4.6	989	1	US-10-821-234-975
18	92.5	4.5	853	1	US-10-821-234-1110
19	91.5	4.5	299	1	US-10-793-626-1888
20	91.5	4.5	636	1	US-10-485-517-170
21	91.5	4.5	1279	1	US-10-793-626-3188
22	91	4.5	574	1	US-10-507-275-7
23	90.5	4.4	296	1	US-10-131-826A-26
24	90.5	4.4	672	1	US-10-689-742-70
25	90.5	4.4	702	1	US-10-510-386-214

26	90	4.4	410	1	US-10-793-626-3258	Sequence 3258, Ap
27	90	4.4	477	1	US-10-793-626-3250	Sequence 3250, Ap
28	90	4.4	919	1	US-10-821-234-1144	Sequence 1144, Ap
29	89.5	4.4	208	1	US-10-793-626-694	Sequence 694, App
30	89.5	4.4	208	1	US-10-793-626-1326	Sequence 1326, Ap
31	89	4.4	756	7	US-11-074-176-202	Sequence 202, App
32	88	4.3	368	1	US-10-689-742-100	Sequence 100, App
33	87.5	4.3	284	1	US-10-821-234-1632	Sequence 1632, Ap
34	87.5	4.3	594	1	US-10-131-826A-10	Sequence 10, Appl
35	87.5	4.3	654	7	US-11-046-668-4	Sequence 4, Appl
36	87.5	4.3	674	1	US-10-501-039-10	Sequence 10, Appl
37	87.5	4.3	745	1	US-10-793-626-1500	Sequence 1500, Ap
38	87	4.3	709	7	US-11-074-176-158	Sequence 158, App
39	87	4.3	747	1	US-10-821-234-1662	Sequence 1662, Ap
40	87	4.3	770	1	US-10-982-545-15	Sequence 15, Appl
41	87	4.3	770	1	US-10-789-273-38	Sequence 38, Appl
42	87	4.3	1586	1	US-10-821-234-901	Sequence 901, App
43	86.5	4.2	352	1	US-10-793-626-216	Sequence 216, App
44	86.5	4.2	1126	1	US-10-485-517-248	Sequence 248, App
45	86	4.2	752	1	US-10-793-626-1138	Sequence 1138, Ap

ALIGNMENTS

RESULT 1

US-10-793-626-1358
; Sequence 1358, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 1358
; LENGTH: 583
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-1358

Query Match	5.8%	Score 118.5;	DB 1;	Length 583;
Best Local Similarity	20.4%	Pred. No. 0.075;		
Matches	87;	Conservative	75;	Mismatches 170; Indels 95; Gaps 17;
Qy	46	MSESHQRQLLASSENPOQFMDYF---SESPR-----NDFLELLRRRFGTKRVH---NNI 93		
Db	120	VNQNGQFIVQLLGWNAEQFQLFILPQGFEPKFLQSNKDKQILRLTFNSERFDEIRHL 179		
Qy	94	VYNEYISHREIHMNATOMETLTDTFKWLGREGKLVDETPKGMVQIYIDRDE-----147		
Db	180	LVENVKQKQVQIENRYTQIENLWNDIDTFNNDELALYKELESSTQDKMIEKFFQFNDYGC 239		
Qy	148	TIRROLELEKKK-KQDLDD-----EETAKFIEQVRRLGEGKEQ-----187		
Db	240	KILKSFEAKNKITKELDDLNLHKYKYNVELSENKKLKAERKIPDDLKKEQNYIDKLKQE 299		
Qy	188	-----VPTFTTEL-SRENDEEKVTFNLSKGACSSGATSSKSTLSPALKTTGSS 236		
Db	300	LKMQESKVLITVTRLSLQSLKDKDELV-----SLHEQSKLNETNYHNEIKGPKQK 350		
Qy	237	----ASVKRESSQSSTQSEKKKKKSLADEIME-----IEEKKRTARTDYWLQPEII 286		
Db	351	LEHLSTRENEITQFNOYLEKNQVFFNQDLKIISYQOKPVEIEIKKL-----YSEYNDLI 406		
Qy	287	VK--IITKLGEKYHKKKAIVKEVIDKYTAVVMIDSG-----DKLKLDQTHLET 334		


```
Qy 131 DETPKGMYIOYIDRDPETIRROLELEKKKKQDLDDEK-----TAKF----- 172
Db 814 D--PK-----IAYTNKLENLSQ-----ENKQHQIDNSEKQIEDITAKLTILAQNDESNMQ 865
Qy 173 ---IEEVRRCLEGEQVPT-FTELSRE-----NDEEKV---TFNLSKGACSSGA 217
Db 866 TANLEKQ-KSTIEQKNQLARLNDLSSQLGQPDQINQLDQVARNYDLRKDAAEQED 924
Qy 218 TSXKSSSTLGS---ALKTIGSSASVRKES-SQSSTOSKEKKKKKSA----- 260
Db 925 YSVKIAKFNSSINQRLTDRDYSLTFFAAIAQAGENNENETRNELAKSVKLHRMSIEDI 984
Qy 261 ----LDIMEIEEKKR-----TARTDYWLQPEIIVKIITKLGEXYHKKKAI 304
Db 985 GPNVLDISIQYEDVKQRYDFNGQONDLLKARD-----LEKSMTELDDEVKTRFKHT 1037
Qy 305 VKEVIDKYTAVVMKIDSGDKLKD-----QTHLETVIPAPGK---RILVNGGYR 351
Db 1038 FDTIAESFQKIFPVVFGGKAKLELTPDNLLETGIEIIAQPQKGLQRLSLSGGER 1095

RESULT 5
US-10-485-517-212
; Sequence 212, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629W0
; CURRENT APPLICATION NUMBER: US/10/485.517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 212
; LENGTH: 1448
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-10-485-517-212

Query Match 5.1%; Score 104.5; DB 1; Length 1448;
Best Local Similarity 20.1%; Pred. No. 2.3;
Matches 58; Conservative 47; Mismatches 92; Indels 91; Gaps 12;

Qy 129 KVDETPKGYIYIDRDPETIRROLELEKKKKQDLDDEKT-----AKFIEQVRRGL-- 181
Db 803 QVDATVFPKQAIEVKAEDTKESID-----QSDQLTAEKTEALAMIKQITDQAKQGITD 858
Qy 182 -----EGKEQVPTTETLSRENDE-----EKVTFNLSKGACSSGATSKSSTL 225
Db 859 ATTAAEVEKAKAQAQLEAFDNIQIDSTEKQKAIETALDQIEAGVNVNADATTEKEAF 918
Qy 226 GPSALKTTGSSASVRKESQSSTOSKEKKKKKSGALDIMEIEEKKRTARTDYWLQPEI 285
Db 919 -TNALEDILSKAT---EIDISQTNAAETATVKSNALEQL-----KAQR-----INFEV 962
Qy 286 IVKIITKLGEXYHKKKAI--VKEVIDKYTAVVMKIDS-----GDK 324
Db 963 -----KKNALAEIREVVNQIIEIKNADADASAKETARTDLGRYDFRADK 1008
Qy 325 LKLDQTH-----LETVIPAGKRLVLNGGYRNEGTL 359
Db 1009 LDKQTQNAEVAELQNVTPATEAIVPQNDPANDTNGNDNDATANS 1056

RESULT 6
US-10-793-626-2964
```

```
; Sequence 2964, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P034800S
; CURRENT APPLICATION NUMBER: US/10/793.626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2964
; LENGTH: 5024
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (5024)
; OTHER INFORMATION: variable amino acid
; US-10-793-626-2964

Query Match 5.0%; Score 101.5; DB 1; Length 5024;
Best Local Similarity 20.7%; Pred. No. 18;
Matches 67; Conservative 60; Mismatches 119; Indels 77; Gaps 16;

Qy 91 NNIVYNEYISHREH-----IHMNATQWETLTDF-TKWLREGGLCK-VDETPKGYI 139
Db 3463 NNLHGQKLAHAKQDAANVINGLIHLNVAQREVMINTNTATTREKVAKLDNA----- 3516
Qy 140 QYIDRDPETIRROLELEKKKKQDLDDEE---KTAKFIEEQRVRRGLEGEDEV----- 188
Db 3517 QALDKAMETLQ---QVVAHKNILNDSKYLNESKY---QOQYDRIAEOQLNQTTPTL 3572
Qy 189 -PTFTELSREN--DEEKVTFNLSKGACSSGATS-----SKSSTLGPALKTIGSSASVK 240
Db 3573 EPYKVIDKONVLANEKILFGAEKLSYDKSNANDEIKHMYLNNAQKQSKDMISHAALR 3632
Qy 241 RKSSQSSTOSKEKKKKKSGALDIMEIEEKKRTARTDYWLQPEIIVKIITKLGEXYHK 300
Db 3633 -----TEVKQLLQAKTLDEAMKSLDKTQVITDITL-----PNYTEASED 3674
Qy 301 KKAIVKEVIDKYTAVVMKIDSGDKLKDQTH--LETVIPAGKRLVLNGGYRNEGTL-- 356
Db 3675 KKEKVDQTVSHAQAIIDIKI-NGSNVSLDQVRQALEQLTQASEN-----LDGQORVEEAKVH 3729
Qy 357 -----LESINEKTFSATI 369
Db 3730 ANQTIDQLTHLSLQOQTAKESV 3752

RESULT 7
US-10-507-275-9
; Sequence 9, Application US/10507275
; Publication No. US20050250166A1
; GENERAL INFORMATION:
; APPLICANT: Masai, Hiseao
; APPLICANT: Tamai, Katsuyuki
; APPLICANT: Medical and Biological Laboratories Co., Ltd.
; APPLICANT: Japan Science and Technology Agency
; APPLICANT: Ginkgo Biomedical Research Institute Co., Ltd.
; TITLE OF INVENTION: Coc7-ASK Kinase Complex, Substrates of the Kinase Complex,
; TITLE OF INVENTION: Specific Antibodies to the Substrates, and Screening Methods
; TITLE OF INVENTION: Using the Same to Screen for Compounds Comprising Cdc7-ASK
; TITLE OF INVENTION: Kinase Inhibitory Ability
; FILE REFERENCE: 082368-001100US
; CURRENT APPLICATION NUMBER: US/10/507.275
; CURRENT FILING DATE: 2004-09-09
; PRIOR APPLICATION NUMBER: JP 2002-067702
; PRIOR FILING DATE: 2002-03-12
```



```
Qy 372 ETGLKGRVVEG---IQYEDISKL 392
Db 401 EVG--IGDISGDPISQLNEMPHL 423

RESULT 10
US-10-793-626-348
; Sequence 348, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P03480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 348
; LENGTH: 752
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-348

Query Match 4.8%; Score 98.5; DB 1; Length 752;
Best Local Similarity 20.9%; Pred. No. 2.7;
Matches 71; Conservative 69; Mismatches 112; Indels 87; Gaps 19;

Qy 123 GREGLCKVDTPKGYI--QVIDRDP-ETIRROL-----ELEK-----KKQDLDEEK 168
Db 174 GIEGKAYETGKGVVRSRDEPLRSGRQLIVTEIPYEVNKSILVKRIDLADKK 233
Qy 169 TAKTIE---EQVRGLE-----GKEQEVPTTFLSRENDEEKVTNLSKGCSSSGATSS 220
Db 234 VDGIVEVRDEYDRTGLRIALKKDANSEIKNLYKNSDLQISYFNFWAIS-----EG 288
Qy 221 KSSTLGPSALTKTGSSASVKRKESSQST---QSKERK-----KKKSALDEIMIE 268
Db 289 RPKLMG---LREIETSYLNHQIEVYVNTNRYDLEQAERKMHVGLMKALSILDEVIALI 345
Qy 269 EEKX-----TARTDYW-LQPEIIVKI-----IT-----KKLGEKYHKKAIVKE---V 308
Db 346 RNSKNKKDKNLVAEYDTEAQAEAIVWLQYRLTNTDIEALKKEHELEBALIKELRNI 405
Qy 309 IDKYTAVVVM1-----DSGDKLKLDOTHL-----ETVIPAPGKRLVILNGGYR 351
Db 406 LDNHEALLAVIKDELNEIKKPKVDRLSTIEAEISEIKIDKEVMVPSSEVILSLTQHG- 464
Qy 352 GNEGTLSEINKEPTGATIVETGPKLGRVVEGIVQVEDIS 390
Db 465 -----IKTSTRSFNASGVTEIGLXGDRL--LKHSYV 496

RESULT 11
US-11-135-855-28
; Sequence 28, Application US/11135855
; Publication No. US20050255557A1
; GENERAL INFORMATION:
; APPLICANT: SMITHKLINE BEECHAM CORPORATION
; APPLICANT: SMITHKLINE BEECHAM P.L.C.
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50013
; CURRENT APPLICATION NUMBER: US/11/135,855
; CURRENT FILING DATE: 2005-05-24
; PRIOR APPLICATION NUMBER: US/10/203,708
; PRIOR FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: PCT/US01/04703
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/182,172
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 29
; LENGTH: 717
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-135-855-29

Query Match 4.7%; Score 95.5; DB 7; Length 717;
Best Local Similarity 24.3%; Pred. No. 4.1;
Matches 51; Conservative 30; Mismatches 64; Indels 65; Gaps 10;

Qy 144 RDPETIRROLELEKKKKQD-----LDDEETAKTFIEQVRR-----GLEGK 184
Db 359 RRDEARRRELEARRRREQESELRLREQK-----EKKERRRRADRGEAERGSGSGD 413
Qy 185 EQEVPPTFELSRENDEEKVTNLSKGCSSSGATSSKSTLGPALKTIIGSSASVKRKE 244
Db 414 E-----LREDDP-----VKGRGKGRGPPSSDSEPEALELREAKKSAKKPQS 459
Qy 245 SQSSTQSKKKKKSALDEIMEIEBEKRTARTDYWLOPEIIVKI-ITKKLGEKY----- 298
```

Db 460 SSTEPARKCQKEK-----RVREKQQAAP-----VKVTRKRSKSGFSDMRK 503
 Qy 299 --HKKAIIVKEIDKYTAVVKN---IDSGB 323
 Db 504 VEKKEPSVEEKQLKHSIKFALKVDSFD 533
 RESULT 13
 US-10-821-234-1313
 ; Sequence 1313, Application US/10821234
 ; Publication No. US20050255114A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Labat, Ivan
 ; APPLICANT: Stache-Crain, Birgit
 ; APPLICANT: Andarmani, Susan
 ; APPLICANT: Tang, Y. Tom
 ; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
 ; FILE REFERENCE: 821A
 ; CURRENT APPLICATION NUMBER: US/10/821,234
 ; CURRENT FILING DATE: 2004-04-07
 ; PRIOR APPLICATION NUMBER: US 60/462,047
 ; PRIOR FILING DATE: 2003-04-07
 ; NUMBER OF SEQ ID NOS: 1704
 ; SOFTWARE: pt_SEQ_genes Version 1.0
 ; SEQ ID NO 1313
 ; LENGTH: 422
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-821-234-1313
 Query Match 4.7%; Score 95; DB 1; Length 422;
 Best Local Similarity 21.0%; Pred. No. 2.3;
 Matches 87; Conservative 59; Mismatches 128; Indels 140; Gaps 19;
 Qy 24 LRMYCQKQCRDENGFKCHMSH-----QRQLLASNPQPFMDYFS 69
 Db 105 LKWH-----QORDPGSKDYSQYENITHLQEQIVDGNMTNAQIILIDNARMAVDNF 158
 Qy 70 EEPNDF-----LELRRRFGTKRVHNNI-----VYNEYISHREHIHMNAQ 111
 Db 159 LKYENHSFKDLEI--EVEGLRRILDNLTIVTTDLQEVGMKXELILMKKHHEQEMEK 216
 Qy 112 WETLFTKWLREGGLCKVDTPKGYIYIDRDPETIRROLEL-EKKKKQDLD--BEK 168
 Db 217 HHVPSDFNVN-----KVDTPREDLIKVL---EDMRQVELIILKKHRDLDTWYKEQ 266
 Qy 169 TAKFIEQVRGLEKEQVPTFELSREND--EEKVTFNLKSGACSSGATSKSSTLG 226
 Db 267 SAAMSQEAASPA-----TVQSRQGDTHLKRFTQALEIDLQTYSTK----- 308
 Qy 227 PSALKTISSASVKRKSSQSTQSKKKKKSALDEIMEIEEKKRTARTDYWLQPEII 286
 Db 309 -SAL-----ENMLSTQARNCKLQDMQEIISHYEEELTQ-----LRHEL- 347
 Qy 287 VKIITKKLGEYHKKAIIVKEIDKYTAVVKNIDSGDKLKDQTHLETVPAPGKRILVL 346
 Db 348 -----ERQNNYQVLLGI-----KTHLEKEI--TTYRRLI-- 375
 Qy 347 NGYGRNGTLESINEKTSATVIETGPKLGRRVEGI-----QYEDISLA 393
 Db 376 -----EGESEGTREESKSMKVSATPKIKAITQETINGRLVLCQVNEIQHA 422
 RESULT 14
 US-10-793-626-1780
 ; Sequence 1780, Application US/10793626
 ; Publication No. US20050255478A1
 ; GENERAL INFORMATION:
 ; APPLICANT: KIMBERLY, WILLIAM JOHN
 ; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
 ; FILE REFERENCE: PU3480US
 ; CURRENT APPLICATION NUMBER: US/10/793,626
 ; CURRENT FILING DATE: 2004-03-04

This Page Blank (uspicio)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:33:44 ; Search time 133.339 Seconds
(without alignments)
1295.011 Million cell updates/sec

Title: US-09-555-529-26
Perfect score: 393
Sequence: 1 MGKSDFTPKAIANRIKSG.....GPKLGRVGEIQYEDISKLA 393

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0
Searched: 2443163 seqs, 439378781 residues
Word size : 0

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A Geneseq 21.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*
9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	393	100.0	393	2	AA23620 A human k
2	393	100.0	393	7	ADJ69659 Human hea
3	167	42.5	391	2	AA23619 Murine ki
4	165	42.0	293	2	AA23618 A huamn t
5	120	30.5	291	2	AA23617 Mouse tru
6	73	18.6	84	4	AA23617 Human col
7	42	10.7	390	8	ADP22452 Sea-squir
8	35	8.9	411	3	AA23619 Arabidops
9	35	8.9	423	3	AA23619 Arabidops
10	35	8.9	423	3	AA23619 Arabidops
11	35	8.9	423	3	AA23619 Arabidops
12	35	8.9	423	3	AA23619 Arabidops
13	35	8.9	423	3	AA23619 Arabidops
14	35	8.9	423	3	AA23619 Arabidops
15	24	6.1	382	3	AA23619 Arabidops
16	21	5.3	302	3	AA23619 Arabidops
17	21	5.3	340	3	AA23619 Arabidops
18	21	5.3	343	3	AA23619 Arabidops
19	18	4.6	34	5	ADK35903 Novel hum
20	18	4.6	38	5	ADK35871 Novel hum
21	15	3.8	15	2	AA23619 Arabidops
22	15	3.8	390	4	AB271416 Drosophil
23	11	2.8	11	2	AA23619 Arabidops
24	8	2.0	8	5	AB23619 Arabidops

25	8	2.0	8	5	AB23619 Arabidops
26	8	2.0	8	7	ADL99055 Human leu
27	8	2.0	8	7	ADL98437 Human leu
28	8	2.0	55	3	AA23619 Arabidops
29	8	2.0	61	2	AA23619 Arabidops
30	8	2.0	61	5	AB23619 Arabidops
31	8	2.0	61	6	AB23619 Arabidops
32	8	2.0	61	7	AD23330 Novel hum
33	8	2.0	61	8	ADH74332 Human sec
34	8	2.0	96	5	AB23619 Arabidops
35	8	2.0	113	3	AA23619 Arabidops
36	8	2.0	119	4	AA23619 Arabidops
37	8	2.0	129	8	ADY13398 Human pol
38	8	2.0	146	4	AB23619 Arabidops
39	8	2.0	173	4	AAU58581 Propionib
40	8	2.0	173	6	AB23619 Arabidops
41	8	2.0	176	2	AA23619 Arabidops
42	8	2.0	178	4	AA23619 Arabidops
43	8	2.0	330	3	AA23619 Arabidops
44	8	2.0	330	3	AA23619 Arabidops
45	8	2.0	351	3	AA23619 Arabidops

ALIGNMENTS

RESULT 1
AA23620
ID AAY23620 standard; protein; 393 AA.
XX
AC AAY23620;
XX
DT 07-SEP-1999 (first entry)
XX
DE A human kin17 protein.
XX
KW Human; kin17 protein; cell proliferation; fertility;
KW hyperproliferative disease; protein interaction; curved DNA;
KW HIV replication; HIV integration; repair enzyme.
XX
OS Homo sapiens.
XX
PN FR2772046-A1.
XX
PD 11-JUN-1999.
XX
PF 09-DEC-1997; 97FR-00015536.
XX
PR 09-DEC-1997; 97FR-00015536.
(COMS) COMMISSARIAT ENERGIE ATOMIQUE.
Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
WPI; 1999-359999/31.
New DNA coding for human kin17 protein - useful for controlling cell proliferation or fertility.
Claim 19; Page 37-38; 69pp; French.
The present sequence represents a human kin17 protein. The mammalian kin17 protein is useful for preparing a medicament for controlling cell proliferation or for controlling fertility. The medicaments can also be used to treat hyperproliferative diseases. Fragments between amino acids 55 and 235 (preferably between amino acids 129 and 228) of a mammalian kin17 protein are useful for regulating the interaction between proteins and curved DNA. The fragment can be used to block replication of HIV or its integration into the human genome or to target repair enzymes to curved DNA sites. Expression vectors for kin17 can be used for controlling cell proliferation
Sequence 393 AA;


```

PN FR2772046-A1.
XX
PD 11-JUN-1999.
XX
PF 09-DEC-1997; 97FR-00015536.
XX
PR 09-DEC-1997; 97FR-00015536.
XX
PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.
XX
PI Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
XX WPI; 1999-359999/31.
XX
PT New DNA coding for human kin17 protein - useful for controlling cell
XX proliferation or fertility.
XX
PS Claim 19; Page 36-37; 69pp; French.
XX
CC The present sequence represents a murine kin17 protein. The mammalian
CC kin17 protein is useful for preparing a medicament for controlling cell
CC proliferation or for controlling fertility. The medicaments can also be
CC used to treat hyperproliferative diseases. Fragments between amino acids
CC 55 and 235 (preferably between amino acids 129 and 228) of a mammalian
CC kin17 protein are useful for regulating the interaction between proteins
CC and curved DNA. The fragment can be used to block replication of HIV or
CC its integration into the human genome or to target repair enzymes to
CC curved DNA sites. Expression vectors for kin17 can be used for
CC controlling cell proliferation
XX
SQ Sequence 391 AA;
Query Match 42.5%; Score 167; DB 2; Length 391;
Best Local Similarity 100.0%; Pred. No. 6.4e-150;
Matches 167; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 21 LOKLRWYCOMKOCORDENGFKCHCMSESHORQLLASENPQQFMDYFSEFRNDFLELL 80
Db 21 LOKLRWYCOMKOCORDENGFKCHCMSESHORQLLASENPQQFMDYFSEFRNDFLELL 80
Qy 81 RRRFGTKRVHNNIVYNEIYISHREIHMNATQWETLDTFTKWLGRGLCKVDETPKGYIQ 140
Db 81 RRRFGTKRVHNNIVYNEIYISHREIHMNATQWETLDTFTKWLGRGLCKVDETPKGYIQ 140
Qy 141 YIDRDPETIRQLLEKXKQDLDDDEKTAKFIEQVRRGLGKEQE 187
Db 141 YIDRDPETIRQLLEKXKQDLDDDEKTAKFIEQVRRGLGKEQE 187
RESULT 4
AAY23618
ID AAY23618 standard; protein; 293 AA.
AC AAY23618;
XX
DT 07-SEP-1999 (first entry)
XX
DE A huamn truncated kin17 protein.
XX
KW kin17 protein; cell proliferation; fertility; hyperproliferative disease;
KW protein interaction; curved DNA; HIV replication; HIV integration;
KW repair enzyme.
XX
OS Homo sapiens.
XX
PN FR2772046-A1.
XX
PD 11-JUN-1999.
XX
PF 09-DEC-1997; 97FR-00015536.
XX
PR 09-DEC-1997; 97FR-00015536.
XX
PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.
XX
PI Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
XX WPI; 1999-359999/31.
XX
PT New DNA coding for human kin17 protein - useful for controlling cell
XX proliferation or fertility.
XX
PS Claim 19; Page 36-37; 69pp; French.
XX
CC The present sequence represents a murine kin17 protein. The mammalian
CC kin17 protein is useful for preparing a medicament for controlling cell
CC proliferation or for controlling fertility. The medicaments can also be
CC used to treat hyperproliferative diseases. Fragments between amino acids
CC 55 and 235 (preferably between amino acids 129 and 228) of a mammalian
CC kin17 protein are useful for regulating the interaction between proteins
CC and curved DNA. The fragment can be used to block replication of HIV or
CC its integration into the human genome or to target repair enzymes to
CC curved DNA sites. Expression vectors for kin17 can be used for
CC controlling cell proliferation
XX
SQ Sequence 391 AA;
Query Match 42.5%; Score 167; DB 2; Length 391;
Best Local Similarity 100.0%; Pred. No. 6.4e-150;
Matches 167; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 21 LOKLRWYCOMKOCORDENGFKCHCMSESHORQLLASENPQQFMDYFSEFRNDFLELL 80
Db 21 LOKLRWYCOMKOCORDENGFKCHCMSESHORQLLASENPQQFMDYFSEFRNDFLELL 80
Qy 81 RRRFGTKRVHNNIVYNEIYISHREIHMNATQWETLDTFTKWLGRGLCKVDETPKGYIQ 140
Db 81 RRRFGTKRVHNNIVYNEIYISHREIHMNATQWETLDTFTKWLGRGLCKVDETPKGYIQ 140
Qy 141 YIDRDPETIRQLLEKXKQDLDDDEKTAKFIEQVRRGLGKEQE 187
Db 141 YIDRDPETIRQLLEKXKQDLDDDEKTAKFIEQVRRGLGKEQE 187
RESULT 4
AAY23618
ID AAY23618 standard; protein; 293 AA.
AC AAY23618;
XX
DT 07-SEP-1999 (first entry)
XX
DE Mouse truncated kin17 protein.
XX
KW kin17 protein; cell proliferation; fertility; hyperproliferative disease;
KW protein interaction; curved DNA; HIV replication; HIV integration;
KW repair enzyme.
XX
OS Mus sp.
XX
PN FR2772046-A1.
XX
PD 11-JUN-1999.
XX
PF 09-DEC-1997; 97FR-00015536.
XX
PR 09-DEC-1997; 97FR-00015536.
XX
PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.
XX
PI Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
XX WPI; 1999-359999/31.
XX
PT New DNA coding for human kin17 protein - useful for controlling cell
XX proliferation or fertility.
XX
PS Claim 15; Page 34-35; 69pp; French.
XX
CC The present sequence represents a truncated human kin17 protein with
CC amino acids 129-228 deleted. The mammalian kin17 protein is useful for
CC preparing a medicament for controlling cell proliferation or for
CC controlling fertility. The medicaments can also be used to treat
CC hyperproliferative diseases. Fragments between amino acids 55 and 235
CC (preferably between amino acids 129 and 228) of a mammalian kin17 protein
CC are useful for regulating the interaction between proteins and curved
CC DNA. The fragment can be used to block replication of HIV or its
CC integration into the human genome or to target repair enzymes to curved
CC DNA sites. Expression vectors for kin17 can be used for controlling cell
CC proliferation
XX
SQ Sequence 293 AA;
Query Match 42.0%; Score 165; DB 2; Length 293;
Best Local Similarity 100.0%; Pred. No. 4e-148;
Matches 165; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 229 ALKTIGSSASVVRKESQSQSTQSKKKKKKSSALDEIMEIEEEKKRTARTDYWLQPEIIVK 288
Db 129 ALKTIGSSASVVRKESQSQSTQSKKKKKKSSALDEIMEIEEEKKRTARTDYWLQPEIIVK 188
Qy 289 IITKLGKGYHKKKAIYKEVIDKYTAVVWMDSGDKLKDQTHLETVPAPGKRILVNG 348
Db 189 IITKLGKGYHKKKAIYKEVIDKYTAVVWMDSGDKLKDQTHLETVPAPGKRILVNG 248
Qy 349 GYRNGEGLTLESINEKTFSAIVTETGPKGRRVEGQYEDISKIA 393
Db 249 GYRNGEGLTLESINEKTFSAIVTETGPKGRRVEGQYEDISKIA 293
RESULT 5
AAY23617
ID AAY23617 standard; protein; 291 AA.
AC AAY23617;
XX
DT 07-SEP-1999 (first entry)
XX
DE Mouse truncated kin17 protein.
XX
KW kin17 protein; cell proliferation; fertility; hyperproliferative disease;
KW protein interaction; curved DNA; HIV replication; HIV integration;
KW repair enzyme.
XX
OS Mus sp.
XX
PN FR2772046-A1.
XX
PD 11-JUN-1999.
XX
PF 09-DEC-1997; 97FR-00015536.
XX
PR 09-DEC-1997; 97FR-00015536.
XX
PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.
XX
PI Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
XX WPI; 1999-359999/31.
XX
PT New DNA coding for human kin17 protein - useful for controlling cell
XX proliferation or fertility.
XX
PS Claim 15; Page 34-35; 69pp; French.
XX
CC The present sequence represents a truncated human kin17 protein with
CC amino acids 129-228 deleted. The mammalian kin17 protein is useful for
CC preparing a medicament for controlling cell proliferation or for
CC controlling fertility. The medicaments can also be used to treat
CC hyperproliferative diseases. Fragments between amino acids 55 and 235
CC (preferably between amino acids 129 and 228) of a mammalian kin17 protein
CC are useful for regulating the interaction between proteins and curved
CC DNA. The fragment can be used to block replication of HIV or its
CC integration into the human genome or to target repair enzymes to curved
CC DNA sites. Expression vectors for kin17 can be used for controlling cell
CC proliferation
XX
SQ Sequence 293 AA;
Query Match 42.0%; Score 165; DB 2; Length 293;
Best Local Similarity 100.0%; Pred. No. 4e-148;
Matches 165; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 229 ALKTIGSSASVVRKESQSQSTQSKKKKKKSSALDEIMEIEEEKKRTARTDYWLQPEIIVK 288
Db 129 ALKTIGSSASVVRKESQSQSTQSKKKKKKSSALDEIMEIEEEKKRTARTDYWLQPEIIVK 188
Qy 289 IITKLGKGYHKKKAIYKEVIDKYTAVVWMDSGDKLKDQTHLETVPAPGKRILVNG 348
Db 189 IITKLGKGYHKKKAIYKEVIDKYTAVVWMDSGDKLKDQTHLETVPAPGKRILVNG 248
Qy 349 GYRNGEGLTLESINEKTFSAIVTETGPKGRRVEGQYEDISKIA 393
Db 249 GYRNGEGLTLESINEKTFSAIVTETGPKGRRVEGQYEDISKIA 293

```

PT proliferation or fertility.
PS Claim 14; Page 33-34; 69pp; French.
XX
CC The present sequence represents a mouse kin17 protein with amino acids
CC 129-228 deleted. The mammalian kin17 protein is useful for preparing a
CC medicament for controlling cell proliferation or for controlling
CC fertility. The medicaments can also be used to treat hyperproliferative
CC diseases. Fragments between amino acids 55 and 235 (preferably between
CC amino acids 129 and 228) of a mammalian kin17 protein are useful for
CC regulating the interaction between proteins and curved DNA. The fragment
CC can be used to block replication of HIV or its integration into the human
CC genome or to target repair enzymes to curved DNA sites. Expression
CC vectors for kin17 can be used for controlling cell proliferation
XX
SQ Sequence 291 AA;
Query Match 30.5%; Score 120; DB 2; Length 291;
Best Local Similarity 100.0%; Pred. No. 2.7e-105;
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 PRAIANRIKSGIQLRWYQCMQKQCRDENGFKCHCMSESHQRLLLASENPQQFMDYF 68
DB 9 PRAIANRIKSGIQLRWYQCMQKQCRDENGFKCHCMSESHQRLLLASENPQQFMDYF 68
QY 69 SEEFNRDLELLRRRGTGKRVHNNIVYNEYISHREHNMNATQWETLDTFTKWLREGLC 128
DB 69 SEEFNRDLELLRRRGTGKRVHNNIVYNEYISHREHNMNATQWETLDTFTKWLREGLC 128
RESULT 6
AAG74690
ID AAG74690 standard; protein; 84 AA.
XX
AC AAG74690;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:5454.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-05026524.
XX
PR 29-SEP-1999; 99US-0157137P.
PR 03-NOV-1999; 99US-0163280P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
XX WPI: 2001-235357/24.
XX
XX N-PSDB; AAH34095.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers.
XX
PS Claim 11; Page 7064-7065; 9803pp; English.
XX
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene therapy
CC and vaccine production. N and P may be used in the prevention, diagnosis
CC and treatment of diseases associated with inappropriate P expression. For
CC example, N and P may be used to treat disorders associated with decreased

CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of P by expressing inactive proteins or to
CC supplement the patient's own production of P. Additionally, N may be used
CC to produce the colon cancer-associated P, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the proteins. N and P
CC can be used in the prevention, diagnosis and treatment of colorectal
CC carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent
CC sequences used in the exemplification of the present invention. N.B.
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC time of publication, meaning no sequences are present for SEQ ID NO:1027
CC to 1052, 7921 and 7922
XX
SQ Sequence 84 AA;
Query Match 18.6%; Score 73; DB 4; Length 84;
Best Local Similarity 100.0%; Pred. No. 4.8e-61;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 321 SGDKLKDQTHLETVIPAPGKRILVNGYRGNEGTLESINEKTSATIVETGPKGRR 380
DB 12 SGDKLKDQTHLETVIPAPGKRILVNGYRGNEGTLESINEKTSATIVETGPKGRR 71
QY 381 VEGIOVEDISKLA 393
DB 72 VEGIOVEDISKLA 84
RESULT 7
ADP22452
ID ADP22452 standard; protein; 390 AA.
XX
AC ADP22452;
XX
DT 12-AUG-2004 (first entry)
XX
DE Sea-squirt (Ciona intestinalis) zinc finger protein #13.
XX
KW sea-squirt; zinc finger protein; gene detection; drug development;
KW zinc finger protein-associated disease.
XX
OS Ciona intestinalis.
XX
PN JP2004057126-A.
XX
PD 26-FEB-2004.
XX
PF 31-JUL-2002; 2002JP-00222484.
XX
PR 31-JUL-2002; 2002JP-00222484.
XX
XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX
XX WPI: 2004-208711/20.
XX
XX N-PSDB; ADP22451.
XX
XX Novel gene encoding zinc finger protein, useful as probe in gene
XX detecting instruments and in development of drug for treating zinc finger
XX protein associated diseases.
XX
PS Claim 1; SEQ ID NO 26; 972pp; Japanese.
XX
XX The invention comprises the amino acid and coding sequences of sea-squirt
XX (Ciona intestinalis) zinc finger proteins. The DNA and protein sequences
XX of the invention are useful in a gene detecting instrument. The DNA and
XX protein sequences of the invention are useful in the development of drugs
XX for the treatment of zinc finger protein-associated diseases. The present
XX amino acid sequence represents a sea-squirt zinc finger protein of the
XX invention.
XX
SQ Sequence 390 AA;
Query Match 10.7%; Score 42; DB 8; Length 390;
Best Local Similarity 100.0%; Pred. No. 6.4e-31;

		Matches	42;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	14	NRKSKGLOKLRWYCOMQCKQCRDENGFKCHCMSESHQROLL 55									
Db	16	NRKSKGLOKLRWYCOMQCKQCRDENGFKCHCMSESHQROLL 57									
RESULT 8											
ID	AAG42576										
ID	AAG42576 standard; protein; 411 AA.										
XX	AC AAG42576;										
XX	DT 18-OCT-2000 (first entry)										
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 53114.										
XX	Arabidopsis thaliana.										
XX	EP1033405-A2.										
XX	06-SEP-2000.										
XX	25-FEB-2000; 2000EP-00301439.										
XX	25-FEB-1999; 99US-0121825P.										
PR	05-MAR-1999;	99US-0123180P.	PR	16-JUN-1999;	99US-0139453P.	PR	17-JUN-1999;	99US-0139454P.	PR	16-JUN-1999;	99US-0139453P.
PR	09-MAR-1999;	99US-0123548P.	PR	17-JUN-1999;	99US-0139454P.	PR	18-JUN-1999;	99US-0139454P.	PR	17-JUN-1999;	99US-0139454P.
PR	23-MAR-1999;	99US-0125788P.	PR	18-JUN-1999;	99US-0139455P.	PR	18-JUN-1999;	99US-0139455P.	PR	18-JUN-1999;	99US-0139455P.
PR	25-MAR-1999;	99US-0126264P.	PR	18-JUN-1999;	99US-0139456P.	PR	18-JUN-1999;	99US-0139456P.	PR	18-JUN-1999;	99US-0139456P.
PR	29-MAR-1999;	99US-0126785P.	PR	18-JUN-1999;	99US-0139457P.	PR	18-JUN-1999;	99US-0139457P.	PR	18-JUN-1999;	99US-0139457P.
PR	01-APR-1999;	99US-0127462P.	PR	18-JUN-1999;	99US-0139458P.	PR	18-JUN-1999;	99US-0139458P.	PR	18-JUN-1999;	99US-0139458P.
PR	06-APR-1999;	99US-0128234P.	PR	18-JUN-1999;	99US-0139459P.	PR	18-JUN-1999;	99US-0139459P.	PR	18-JUN-1999;	99US-0139459P.
PR	08-APR-1999;	99US-0128714P.	PR	18-JUN-1999;	99US-0139460P.	PR	18-JUN-1999;	99US-0139460P.	PR	18-JUN-1999;	99US-0139460P.
PR	16-APR-1999;	99US-0129845P.	PR	18-JUN-1999;	99US-0139461P.	PR	18-JUN-1999;	99US-0139461P.	PR	18-JUN-1999;	99US-0139461P.
PR	19-APR-1999;	99US-0130077P.	PR	18-JUN-1999;	99US-0139462P.	PR	18-JUN-1999;	99US-0139462P.	PR	18-JUN-1999;	99US-0139462P.
PR	21-APR-1999;	99US-0130449P.	PR	18-JUN-1999;	99US-0139463P.	PR	18-JUN-1999;	99US-0139463P.	PR	18-JUN-1999;	99US-0139463P.
PR	23-APR-1999;	99US-0130510P.	PR	18-JUN-1999;	99US-0139464P.	PR	18-JUN-1999;	99US-0139464P.	PR	18-JUN-1999;	99US-0139464P.
PR	28-APR-1999;	99US-0130891P.	PR	18-JUN-1999;	99US-0139465P.	PR	18-JUN-1999;	99US-0139465P.	PR	18-JUN-1999;	99US-0139465P.
PR	30-APR-1999;	99US-0131449P.	PR	18-JUN-1999;	99US-0139466P.	PR	18-JUN-1999;	99US-0139466P.	PR	18-JUN-1999;	99US-0139466P.
PR	30-APR-1999;	99US-0132048P.	PR	18-JUN-1999;	99US-0139467P.	PR	18-JUN-1999;	99US-0139467P.	PR	18-JUN-1999;	99US-0139467P.
PR	04-MAY-1999;	99US-0132484P.	PR	18-JUN-1999;	99US-0139468P.	PR	18-JUN-1999;	99US-0139468P.	PR	18-JUN-1999;	99US-0139468P.
PR	05-MAY-1999;	99US-0132485P.	PR	18-JUN-1999;	99US-0139469P.	PR	18-JUN-1999;	99US-0139469P.	PR	18-JUN-1999;	99US-0139469P.
PR	06-MAY-1999;	99US-0132486P.	PR	18-JUN-1999;	99US-0139470P.	PR	18-JUN-1999;	99US-0139470P.	PR	18-JUN-1999;	99US-0139470P.
PR	07-MAY-1999;	99US-0132487P.	PR	18-JUN-1999;	99US-0139471P.	PR	18-JUN-1999;	99US-0139471P.	PR	18-JUN-1999;	99US-0139471P.
PR	11-MAY-1999;	99US-0132863P.	PR	18-JUN-1999;	99US-0139472P.	PR	18-JUN-1999;	99US-0139472P.	PR	18-JUN-1999;	99US-0139472P.
PR	14-MAY-1999;	99US-0134218P.	PR	18-JUN-1999;	99US-0139473P.	PR	18-JUN-1999;	99US-0139473P.	PR	18-JUN-1999;	99US-0139473P.
PR	14-MAY-1999;	99US-0134219P.	PR	18-JUN-1999;	99US-0139474P.	PR	18-JUN-1999;	99US-0139474P.	PR	18-JUN-1999;	99US-0139474P.
PR	14-MAY-1999;	99US-0134221P.	PR	18-JUN-1999;	99US-0139475P.	PR	18-JUN-1999;	99US-0139475P.	PR	18-JUN-1999;	99US-0139475P.
PR	14-MAY-1999;	99US-0134370P.	PR	18-JUN-1999;	99US-0139476P.	PR	18-JUN-1999;	99US-0139476P.	PR	18-JUN-1999;	99US-0139476P.
PR	18-MAY-1999;	99US-0134768P.	PR	18-JUN-1999;	99US-0139477P.	PR	18-JUN-1999;	99US-0139477P.	PR	18-JUN-1999;	99US-0139477P.
PR	19-MAY-1999;	99US-0134941P.	PR	18-JUN-1999;	99US-0139478P.	PR	18-JUN-1999;	99US-0139478P.	PR	18-JUN-1999;	99US-0139478P.
PR	20-MAY-1999;	99US-0135124P.	PR	18-JUN-1999;	99US-0139479P.	PR	18-JUN-1999;	99US-0139479P.	PR	18-JUN-1999;	99US-0139479P.
PR	21-MAY-1999;	99US-0135353P.	PR	18-JUN-1999;	99US-0139480P.	PR	18-JUN-1999;	99US-0139480P.	PR	18-JUN-1999;	99US-0139480P.
PR	24-MAY-1999;	99US-0135629P.	PR	18-JUN-1999;	99US-0139481P.	PR	18-JUN-1999;	99US-0139481P.	PR	18-JUN-1999;	99US-0139481P.
PR	25-MAY-1999;	99US-0136021P.	PR	18-JUN-1999;	99US-0139482P.	PR	18-JUN-1999;	99US-0139482P.	PR	18-JUN-1999;	99US-0139482P.
PR	27-MAY-1999;	99US-0136392P.	PR	18-JUN-1999;	99US-0139483P.	PR	18-JUN-1999;	99US-0139483P.	PR	18-JUN-1999;	99US-0139483P.
PR	28-MAY-1999;	99US-0136782P.	PR	18-JUN-1999;	99US-0139484P.	PR	18-JUN-1999;	99US-0139484P.	PR	18-JUN-1999;	99US-0139484P.
PR	01-JUN-1999;	99US-0137222P.	PR	18-JUN-1999;	99US-0139485P.	PR	18-JUN-1999;	99US-0139485P.	PR	18-JUN-1999;	99US-0139485P.
PR	03-JUN-1999;	99US-0137528P.	PR	18-JUN-1999;	99US-0139486P.	PR	18-JUN-1999;	99US-0139486P.	PR	18-JUN-1999;	99US-0139486P.
PR	04-JUN-1999;	99US-0137502P.	PR	18-JUN-1999;	99US-0139487P.	PR	18-JUN-1999;	99US-0139487P.	PR	18-JUN-1999;	99US-0139487P.
PR	07-JUN-1999;	99US-0137724P.	PR	18-JUN-1999;	99US-0139488P.	PR	18-JUN-1999;	99US-0139488P.	PR	18-JUN-1999;	99US-0139488P.
PR	08-JUN-1999;	99US-0138094P.	PR	18-JUN-1999;	99US-0139489P.	PR	18-JUN-1999;	99US-0139489P.	PR	18-JUN-1999;	99US-0139489P.
PR	10-JUN-1999;	99US-0138540P.	PR	18-JUN-1999;	99US-0139490P.	PR	18-JUN-1999;	99US-0139490P.	PR	18-JUN-1999;	99US-0139490P.
PR	10-JUN-1999;	99US-0138847P.	PR	18-JUN-1999;	99US-0139491P.	PR	18-JUN-1999;	99US-0139491P.	PR	18-JUN-1999;	99US-0139491P.
PR	14-JUN-1999;	99US-0139119P.	PR	18-JUN-1999;	99US-0139492P.	PR	18-JUN-1999;	99US-0139492P.	PR	18-JUN-1999;	99US-0139492P.
PR	16-JUN-1999;	99US-0139452P.	PR	18-JUN-1999;	99US-0139493P.	PR	18-JUN-1999;	99US-0139493P.	PR	18-JUN-1999;	99US-0139493P.

```
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148584P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match      8.9%; Score 35; DB 3; Length 411;
Best Local Similarity 100.0%; Pred. No. 3.1e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 KGLQKLRWYCMQCKQCRDENGFKCHCMSESQRQ 53
Db 19 KGLQKLRWYCMQCKQCRDENGFKCHCMSESQRQ 53
```

```
RESULT 9
AAB03064
ID AAB03064 standard; protein; 423 AA.
XX
AC AAB03064;
XX
DT 27-SEP-2000 (first entry)
XX
DE Maize KIN17 orthologue, ZmKINH-2.
XX
KW ZmKINH-2; KIN17 orthologue; maize; zinc finger protein; RecA homologue;
KW nuclear localisation; nonhomologous recombination;
KW illegitimate recombination; double stranded DNA binding; curved DNA;
KW homologous gene targeting; transgenic plant.
XX
OS Zea mays.
XX
PN WO200024900-A1.
XX
PD 04-MAY-2000.
XX
PF 06-OCT-1999; 99WO-US023280.
XX
PR 27-OCT-1998; 98US-0105802P.
XX
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Mahajan PB;
XX
DR WPI; 2000-350741/30.
DR N-PSDB; AAA52590.
XX
PT Nucleic acids encoding maize KIN17 orthologue proteins useful for
PT preventing illegitimate recombination in cells.
XX
PS Claim 11; Page 64-66; 84pp; English.
XX
CC This sequence represents the maize KIN17 orthologue ZmKINH-2. The
CC invention relates to maize KIN17 orthologues (AAB03063-B03065), nucleic
CC acids encoding them (AAA52589-A52591), and expression vectors, transgenic
CC plants and plant seeds comprising nucleotides encoding maize KIN17
CC orthologues. KIN17 has, until now, been found only in animal (avian,
CC rodent and human) cells, this invention being the first report describing
CC the presence of KIN17 in plants. Murine KIN17 was found to have
CC significant homology to Escherichia coli RecA protein, and contains a
CC zinc finger motif and a nuclear localisation signal. KIN17 binds double-
CC stranded DNA, preferentially binding to curved DNA, and forms
CC intranuclear foci on overexpression in mammalian cells. It is also
CC induced on exposure to gamma or ultraviolet radiation. These findings
CC indicate that KIN17 plays a role in non-homologous (illegitimate)
CC recombination, which occurs at higher rates among higher eukaryotes,
CC particularly plants. Illegitimate recombination in plants is a major
CC impediment to the generation of transgenic crops such as maize. Maize
CC KIN17 orthologue nucleic acid sequences may be used to generate
CC transgenic plants. The transgenic plants generated can be monocots or
CC dicots and are particularly maize, soybean, sunflower, sorghum, canola,
CC wheat, alfalfa, cotton, rice, barley and millet. In particular, antisense
CC sequences may be used to reduce KIN17 levels in embryogenic callus or
CC embryo cells, thereby reducing the amount of non-homologous recombination
CC and enhancing homologous gene targeting
XX
SQ Sequence 423 AA;

Query Match      8.9%; Score 35; DB 3; Length 423;
Best Local Similarity 100.0%; Pred. No. 3.2e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 KGLQKLRWYCMQCKQCRDENGFKCHCMSESQRQ 53
Db 19 KGLQKLRWYCMQCKQCRDENGFKCHCMSESQRQ 53

RESULT 10
```


AAB03065
 ID AAB03065 standard; protein; 423 AA.
 AC AAB03065;
 XX
 XX 27-SEP-2000 (first entry)
 DT
 XX
 XX Maize KIN17 orthologue, ZmKINH-3.
 DE
 XX
 XX ZmKINH-3; KIN17 orthologue; maize; zinc finger protein; RecA homologue;
 KW nuclear localisation; nonhomologous recombination;
 KW illegitimate recombination; double stranded DNA binding; curved DNA;
 KW homologous gene targeting; transgenic plant.
 XX
 OS Zea mays.
 XX
 XX WO200024900-A1.
 PN
 XX
 XX 04-MAY-2000.
 PD
 XX
 XX 06-OCT-1999; 99WO-US0233280.
 PF
 XX
 XX 27-OCT-1998; 98US-0105802P.
 PR
 XX
 XX (PION-) PIONEER HI-BRED INT INC.
 PA
 XX
 XX Mahajan PB;
 PI
 XX
 XX WPI; 2000-350741/30.
 DR
 XX
 XX N-PSDB; AAA52591.
 DR
 XX
 XX Nucleic acids encoding maize KIN17 orthologue proteins useful for
 PT preventing illegitimate recombination in cells.
 FT
 XX
 XX Claim 11; Page 64-66; 84pp; English.
 PS
 XX
 XX This sequence represents the maize KIN17 orthologue ZmKINH-3. The
 CC invention relates to maize KIN17 orthologues (AAB03063-B03065), nucleic
 CC acids encoding them (AAA52589-A52591), and expression vectors, transgenic
 CC plants and plant seeds comprising nucleotides encoding maize KIN17
 CC orthologues. KIN17 has, until now, been found only in animal (avian,
 CC rodent and human) cells, this invention being the first report describing
 CC the presence of KIN17 in plants. Murine KIN17 was found to have
 CC significant homology to Escherichia coli RecA protein, and contains a
 CC zinc finger motif and a nuclear localisation signal. KIN17 binds double-
 CC stranded DNA, preferentially binding to curved DNA, and forms
 CC intranuclear foci on overexpression in mammalian cells. It is also
 CC induced on exposure to gamma or ultraviolet radiation. These findings
 CC indicate that KIN17 plays a role in non-homologous (illegitimate)
 CC recombination, which occurs at higher rates among higher eukaryotes,
 CC particularly plants. Illegitimate recombination in plants is a major
 CC impediment to the generation of transgenic crops such as maize. Maize
 CC KIN17 orthologue nucleic acid sequences may be used to generate
 CC transgenic plants. The transgenic plants generated can be monocots or
 CC dicots and are particularly maize, soybean, sunflower, sorghum, canola,
 CC wheat, alfalfa, cotton, rice, barley and millet. In particular, antisense
 CC sequences may be used to reduce KIN17 levels in embryogenic callus or
 CC embryo cells, thereby reducing the amount of non-homologous recombination
 CC and enhancing homologous gene targeting
 XX
 XX Sequence 423 AA;
 SQ
 Query Match 8.9%; Score 35; DB 3; Length 423;
 Best Local Similarity 100.0%; Pred. No. 3.2e-24;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 19 KGLQKLRYCQMCQKQCDENGFKCHCMSESHQ 53
 DB 19 KGLQKLRYCQMCQKQCDENGFKCHCMSESHQ 53
 RESULT 11
 ADT58422

ADT58422 standard; protein; 423 AA.
 ADT58422;
 13-JAN-2005 (first entry)
 Plant polypeptide, SEQ ID 8499.
 plant; transgenic; cold tolerance; growth rate; drought tolerance;
 disease resistance; galactomannan production; plant growth regulator;
 heat tolerance; herbicide tolerance; lignin production;
 extreme osmotic condition tolerance; pathogens resistance;
 pest resistance; yield improvement; seed oil yield; seed protein yield.
 Viridiplantae.
 US2004216190-A1.
 28-OCT-2004.
 18-DEC-2003; 2003US-00739930.
 28-APR-2003; 2003US-00424599.
 28-APR-2003; 2003US-00425115.
 (KOVA/) KOVALIC D K.
 Kovalic DK;
 WPI; 2004-757369/74.
 New recombinant DNA constructs useful in the field of biochemistry and
 genetics, and in particular for producing transgenic plants with improved
 biological characteristics.
 Claim 2; SEQ ID NO 8499; 14pp; English.
 The invention relates a recombinant DNA construct comprising a
 polynucleotide having any of 5544 nucleotide sequences (CDNAS SEQ ID NO:
 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences
 (SEQ ID NO: 5545-11088). The CDNAS and proteins are from corn, soybean,
 Arabidopsis, wheat and rape but the specification does not indicate which
 sequences is derived from which organism. Also included is a method of
 producing a plant having an improved property, comprising transforming a
 plant with a recombinant DNA construct comprising a promoter region
 functional in a plant cell operably joined to a polynucleotide encoding a
 polypeptide associated with the property, and growing the transformed
 plant. The property is selected from improving plant cold tolerance, for
 manipulating growth rate in plant cells by modification of the cell cycle
 pathway, for improving plant drought tolerance, for providing increased
 resistance to plant disease, for galactomannan production, for production
 of plant growth regulators, for improving plant heat tolerance, for
 improving plant tolerance to herbicides, for increasing the rate of
 homologous recombination in plants, for lignin production, for improving
 plant tolerance to extreme osmotic conditions, for improving plant
 tolerance to pathogens or pests, for yield improvement by modification of
 photosynthesis, for modifying seed oil yield and/or content, for
 modifying seed protein yield and/or content, for yield improvement by
 modification of carbohydrate, nitrogen or phosphorus use and/or uptake
 and for yield improvement by providing improved plant growth and
 development under at least one stress condition. The polynucleotide may
 also encode a plant transcription factor. The methods and compositions of
 the present invention are useful in the field of biochemistry and
 genetics, in particular for producing transgenic plants with improved
 biological characteristics such as increased yield, improved nitrogen
 flow, increasing plant tolerance to cold or heat, improving plant
 tolerance to extreme osmotic and drought conditions, and improving plant
 tolerance to plant pests or pathogens. They can also be used in physical
 arrays of molecules, plant breeding markers, computer-based storage and
 analysis systems. The present sequence is one of the 5544 plant protein
 sequences of the invention. Note: The sequence data for this patent did
 not form part of the printed specification, but was obtained in
 electronic format directly from USPTO at

CC or by providing improved plant growth and development under at least one
 CC stress condition or for modifying seed oil or protein yield and/or
 CC content. This is the amino acid sequence of a plant full length insert
 CC polypeptide that can be used in the recombinant DNA construct of the
 CC invention.

XX SQ Sequence 437 AA;

Query Match 8.9%; Score 35; DB 8; Length 437;
 Best Local Similarity 100.0%; Pred. No. 3.3e-24;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRYWCQMCQKQCRDENGFKCHCMSESHORQ 53
 |||||
 Db 32 KGLQKLRYWCQMCQKQCRDENGFKCHCMSESHORQ 66
 |||||

RESULT 14

ADX78817
 ID ADX78817 standard; protein; 437 AA.

XX AC ADX78817;

XX DT 21-APR-2005 (first entry)

XX DE Plant full length insert polypeptide seqid 48183.

XX KW plant protectant; plant growth regulant; gene therapy; plant;
 KW recombinant DNA construct; physical array; plant breeding marker;
 KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
 KW extreme osmotic condition; pathogen tolerance; pest tolerance;
 KW growth rate; cell cycle pathway; disease resistance;
 KW galactomannan production; lignin production; plant growth regulator;
 KW yield; plant growth; plant development; seed oil; protein yield;
 KW protein content.

XX OS Unidentified.

XX PN US2004034888-A1.

XX PD 19-FEB-2004.

XX PF 28-APR-2003; 2003US-00425114.

XX PR 06-MAY-1999; 99US-00304517.

XX PR 05-NOV-2001; 2001US-00985678.

XX PA (LIU/) LIU J.

XX PA (ZHOU/) ZHOU Y.

XX PA (KOVA/) KOVALIC D K.

XX PA (SCRE/) SCREEN S E.

XX PA (TAB/) TABASKA J E.

XX PA (CAOV/) CAO Y.

XX PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;

XX DR WPI; 2004-180133/17.

XX PT New recombinant DNA construct, useful for improving plant tolerance to
 PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
 PT pests, for conferring increased resistance to plant disease, or for
 PT improving yield.

XX PS Claim 1; SEQ ID NO 48183; 15pp; English.

XX CC The invention describes a recombinant DNA construct comprising a
 CC polynucleotide consisting of a sequence encoding an amino acid sequence
 CC available in electronic form from the US patent office at
 CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
 CC of the invention are also useful in physical arrays of molecules and as
 CC plant breeding markers. The recombinant DNA construct is useful for
 CC improving plant tolerance to cold, heat, drought, herbicides, extreme
 CC osmotic conditions, pathogens or pests, for manipulating growth rate in

CC plant cells by modification of the cell cycle pathway, for conferring
 CC increased resistance to plant disease, for producing galactomannan,
 CC lignin or plant growth regulators, for increasing the rate of homologous
 CC recombination in plants, for improving yield by modification of
 CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
 CC or by providing improved plant growth and development under at least one
 CC stress condition or for modifying seed oil or protein yield and/or
 CC content. This is the amino acid sequence of a plant full length insert
 CC polypeptide that can be used in the recombinant DNA construct of the
 CC invention.

XX SQ Sequence 437 AA;

Query Match 8.9%; Score 35; DB 8; Length 437;
 Best Local Similarity 100.0%; Pred. No. 3.3e-24;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRYWCQMCQKQCRDENGFKCHCMSESHORQ 53
 |||||
 Db 32 KGLQKLRYWCQMCQKQCRDENGFKCHCMSESHORQ 66
 |||||

RESULT 15

AAG42577

ID AAG42577 standard; protein; 382 AA.

XX AC AAG42577;

XX DT 19-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 53115.

XX KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-00301439.

XX PR 25-FEB-1999; 99US-0121825P.

XX PR 03-MAR-1999; 99US-0123380P.

XX PR 09-MAR-1999; 99US-0123548P.

XX PR 23-MAR-1999; 99US-0125788P.

XX PR 25-MAR-1999; 99US-0126264P.

XX PR 29-MAR-1999; 99US-0126785P.

XX PR 01-APR-1999; 99US-0127462P.

XX PR 06-APR-1999; 99US-0128234P.

XX PR 08-APR-1999; 99US-0128714P.

XX PR 16-APR-1999; 99US-0129845P.

XX PR 19-APR-1999; 99US-0130077P.

XX PR 21-APR-1999; 99US-0130449P.

XX PR 23-APR-1999; 99US-0130510P.

XX PR 28-APR-1999; 99US-0130891P.

XX PR 30-APR-1999; 99US-0131449P.

XX PR 30-APR-1999; 99US-0132048P.

XX PR 04-MAY-1999; 99US-0132407P.

XX PR 05-MAY-1999; 99US-0132484P.

XX PR 06-MAY-1999; 99US-0132485P.

XX PR 06-MAY-1999; 99US-0132486P.

XX PR 07-MAY-1999; 99US-0132487P.

XX PR 11-MAY-1999; 99US-0132863P.

XX PR 14-MAY-1999; 99US-0134256P.

XX PR 14-MAY-1999; 99US-0134218P.

XX PR 14-MAY-1999; 99US-0134219P.

XX PR 14-MAY-1999; 99US-0134221P.

XX PR 14-MAY-1999; 99US-0134370P.

XX PR 18-MAY-1999; 99US-0134768P.

XX PR 19-MAY-1999; 99US-0134941P.

PR 20-MAY-1999;	99US-0135124P.	PR 02-AUG-1999;	99US-0146386P.
PR 21-MAY-1999;	99US-0135353P.	PR 02-AUG-1999;	99US-0146388P.
PR 24-MAY-1999;	99US-0135629P.	PR 02-AUG-1999;	99US-0146389P.
PR 25-MAY-1999;	99US-0136021P.	PR 03-AUG-1999;	99US-0147038P.
PR 27-MAY-1999;	99US-0136392P.	PR 04-AUG-1999;	99US-0147204P.
PR 28-MAY-1999;	99US-0136782P.	PR 04-AUG-1999;	99US-0147302P.
PR 01-JUN-1999;	99US-0137222P.	PR 05-AUG-1999;	99US-0147192P.
PR 03-JUN-1999;	99US-0137528P.	PR 05-AUG-1999;	99US-0147260P.
PR 04-JUN-1999;	99US-0137502P.	PR 06-AUG-1999;	99US-0147303P.
PR 07-JUN-1999;	99US-0137724P.	PR 06-AUG-1999;	99US-0147416P.
PR 08-JUN-1999;	99US-0138094P.	PR 09-AUG-1999;	99US-0147493P.
PR 10-JUN-1999;	99US-0138540P.	PR 09-AUG-1999;	99US-0147935P.
PR 10-JUN-1999;	99US-0138847P.	PR 10-AUG-1999;	99US-0148171P.
PR 14-JUN-1999;	99US-0139119P.	PR 10-AUG-1999;	99US-0148315P.
PR 16-JUN-1999;	99US-0139452P.	PR 12-AUG-1999;	99US-0148341P.
PR 16-JUN-1999;	99US-0139453P.	PR 13-AUG-1999;	99US-0148565P.
PR 17-JUN-1999;	99US-0139492P.	PR 13-AUG-1999;	99US-0148684P.
PR 18-JUN-1999;	99US-0139454P.	PR 16-AUG-1999;	99US-0149368P.
PR 18-JUN-1999;	99US-0139455P.	PR 17-AUG-1999;	99US-0149175P.
PR 18-JUN-1999;	99US-0139456P.	PR 18-AUG-1999;	99US-0149426P.
PR 18-JUN-1999;	99US-0139457P.	PR 20-AUG-1999;	99US-0149722P.
PR 18-JUN-1999;	99US-0139458P.	PR 20-AUG-1999;	99US-0149723P.
PR 18-JUN-1999;	99US-0139459P.	PR 20-AUG-1999;	99US-0149929P.
PR 18-JUN-1999;	99US-0139460P.	PR 23-AUG-1999;	99US-0149902P.
PR 18-JUN-1999;	99US-0139461P.	PR 23-AUG-1999;	99US-0149930P.
PR 18-JUN-1999;	99US-0139462P.	PR 25-AUG-1999;	99US-0150566P.
PR 18-JUN-1999;	99US-0139463P.	PR 26-AUG-1999;	99US-0150884P.
PR 18-JUN-1999;	99US-0139750P.	PR 27-AUG-1999;	99US-0151065P.
PR 18-JUN-1999;	99US-0139817P.	PR 27-AUG-1999;	99US-0151066P.
PR 21-JUN-1999;	99US-0139817P.	PR 30-AUG-1999;	99US-0151080P.
PR 22-JUN-1999;	99US-0139899P.	PR 31-AUG-1999;	99US-0151303P.
PR 23-JUN-1999;	99US-0140353P.	PR 01-SEP-1999;	99US-0151438P.
PR 23-JUN-1999;	99US-0140354P.	PR 01-SEP-1999;	99US-0151930P.
PR 24-JUN-1999;	99US-0140695P.	PR 07-SEP-1999;	99US-0152363P.
PR 28-JUN-1999;	99US-0140823P.	PR 10-SEP-1999;	99US-0153070P.
PR 29-JUN-1999;	99US-0140991P.	PR 13-SEP-1999;	99US-0153758P.
PR 30-JUN-1999;	99US-0141287P.	PR 15-SEP-1999;	99US-0154018P.
PR 01-JUL-1999;	99US-0141842P.	PR 16-SEP-1999;	99US-0154039P.
PR 01-JUL-1999;	99US-0142154P.	PR 20-SEP-1999;	99US-0154779P.
PR 02-JUL-1999;	99US-0142055P.	PR 22-SEP-1999;	99US-0155139P.
PR 06-JUL-1999;	99US-0142390P.	PR 23-SEP-1999;	99US-0155486P.
PR 08-JUL-1999;	99US-0142803P.	PR 24-SEP-1999;	99US-0155659P.
PR 09-JUL-1999;	99US-0142920P.	PR 28-SEP-1999;	99US-0156458P.
PR 12-JUL-1999;	99US-0142977P.	PR 29-SEP-1999;	99US-0156598P.
PR 13-JUL-1999;	99US-0143542P.	PR 04-OCT-1999;	99US-0157117P.
PR 14-JUL-1999;	99US-0143624P.	PR 05-OCT-1999;	99US-0157533P.
PR 15-JUL-1999;	99US-0144005P.	PR 06-OCT-1999;	99US-0157865P.
PR 16-JUL-1999;	99US-0144085P.	PR 07-OCT-1999;	99US-0158029P.
PR 16-JUL-1999;	99US-0144086P.	PR 08-OCT-1999;	99US-0158232P.
PR 19-JUL-1999;	99US-0144325P.	PR 12-OCT-1999;	99US-0158369P.
PR 19-JUL-1999;	99US-0144331P.	PR 13-OCT-1999;	99US-0159293P.
PR 19-JUL-1999;	99US-0144332P.	PR 13-OCT-1999;	99US-0159294P.
PR 19-JUL-1999;	99US-0144333P.	PR 13-OCT-1999;	99US-0159295P.
PR 19-JUL-1999;	99US-0144334P.	PR 14-OCT-1999;	99US-0159329P.
PR 19-JUL-1999;	99US-0144335P.	PR 14-OCT-1999;	99US-0159330P.
PR 20-JUL-1999;	99US-0144352P.	PR 14-OCT-1999;	99US-0159331P.
PR 20-JUL-1999;	99US-0144632P.	PR 14-OCT-1999;	99US-0159637P.
PR 20-JUL-1999;	99US-0144884P.	PR 14-OCT-1999;	99US-0159638P.
PR 21-JUL-1999;	99US-0144814P.	PR 18-OCT-1999;	99US-0159584P.
PR 21-JUL-1999;	99US-0145086P.	PR 21-OCT-1999;	99US-0160741P.
PR 21-JUL-1999;	99US-0145088P.	PR 21-OCT-1999;	99US-0160767P.
PR 22-JUL-1999;	99US-0145089P.	PR 21-OCT-1999;	99US-0160768P.
PR 22-JUL-1999;	99US-0145087P.	PR 21-OCT-1999;	99US-0160770P.
PR 22-JUL-1999;	99US-0145089P.	PR 21-OCT-1999;	99US-0160814P.
PR 22-JUL-1999;	99US-0145192P.	PR 21-OCT-1999;	99US-0160815P.
PR 23-JUL-1999;	99US-0145145P.	PR 22-OCT-1999;	99US-0160980P.
PR 23-JUL-1999;	99US-0145218P.	PR 22-OCT-1999;	99US-0160981P.
PR 23-JUL-1999;	99US-0145224P.	PR 22-OCT-1999;	99US-0160989P.
PR 26-JUL-1999;	99US-0145276P.	PR 25-OCT-1999;	99US-0161404P.
PR 27-JUL-1999;	99US-0145913P.	PR 25-OCT-1999;	99US-0161405P.
PR 27-JUL-1999;	99US-0145918P.	PR 25-OCT-1999;	99US-0161406P.
PR 27-JUL-1999;	99US-0145919P.	PR 26-OCT-1999;	99US-0161359P.
PR 28-JUL-1999;	99US-0145951P.	PR 26-OCT-1999;	99US-0161360P.

This Page Blank (uspic)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:34:39 ; Search time 154.894 Seconds
(without alignments)
1790.078 Million cell updates/sec

Title: US-09-555-529-26
Perfect score: 393
Sequence: 1 MGKDFLTPKAIANRIKSG.....GPLKGRVBEQIYEDISKLA 393

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 2166443 seqs, 705528306 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Uniprot 05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	393	100.0	393	060870_HUMAN	060870 homo sapien
2	179	45.5	195	Q9CV58_MOUSE	Q9CV58 mus musculus
3	179	45.5	391	Q8K339_MOUSE	Q8K339 mus musculus
4	65	16.5	389	Q6GL97_XENTR	Q6GL97 xenopus tro
5	50	12.7	387	Q68F56_XENLA	Q68F56 xenopus lae
6	48	12.2	427	Q4S5G2_TETNG	Q4S5G2 tetraodon n
7	43	10.9	383	Q5RH25_BRARE	Q5RH25 brachydanio
8	41	10.4	269	Q4TDV9_TETNG	Q4TDV9 tetraodon n
9	35	8.9	402	Q9AY87_ORYSA	Q9AY87 oryza sativ
10	35	8.9	411	Q9ZVU5_ARATH	Q9ZVU5 arabidopsis
11	35	8.9	430	Q75LU5_ORYSA	Q75LU5 oryza sativ
12	29	7.4	88	Q4TDW0_TETNG	Q4TDW0 tetraodon n
13	15	3.8	178	Q6X190_DROVA	Q6X190 drosophila
14	15	3.8	244	Q76926_DROME	Q76926 drosophila
15	15	3.8	390	Q8SXR2_DROME	Q8SXR2 drosophila
16	15	3.8	390	Q9VPH4_DROME	Q9VPH4 drosophila
17	14	3.6	263	Q5CFM3_CRYHO	Q5CFM3 cryptospori
18	14	3.6	265	Q5CYD1_CRYPV	Q5CYD1 cryptospori
19	13	3.3	387	Q7PGN1_ANOGA	Q7PGN1 anopheles g
20	12	3.1	121	Q4XND8_PLACH	Q4XND8 plasmodium
21	12	3.1	239	Q6CXJ7_KLUULA	Q6CXJ7 kluyveromyc
22	12	3.1	378	Q4N9K7_THEPA	Q4N9K7 theileria p
23	12	3.1	412	Q4UG63_THEAN	Q4UG63 theileria a
24	12	3.1	441	Q4XDV6_PLACH	Q4XDV6 plasmodium
25	12	3.1	442	Q8IKG2_PLAF7	Q8IKG2 plasmodium
26	12	3.1	445	Q55D16_DICDI	Q55D16 dictyosteli
27	12	3.1	445	Q4ZD77_PLABE	Q4ZD77 plasmodium
28	12	3.1	457	Q7RD49_PLAYO	Q7RD49 plasmodium
29	11	2.8	238	Q51F62_ENTHI	Q51F62 entamoeba h
30	10	2.5	397	Q60L39_CAEBR	Q60L39 caenorhabdi
31	10	2.5	404	Q9XWF2_CAEEL	Q9XWF2 caenorhabdi

RESULT 1					
O60870_HUMAN					
ID	O60870_HUMAN	PRELIMINARY;	PRT:	393	AA.
AC	O60870;				
DT	01-AUG-1998	(TREMBlrel. 07, Created)			
DT	01-JAN-1999	(TREMBlrel. 09, Last sequence update)			
DT	13-SEP-2005	(TREMBlrel. 31, Last annotation update)			
DE	Kin17 protein (HsKin17 protein) (KIN, antigenic determinant of recA protein homolog) (Mouse).				
DE	Names:KIN; Synonyms=Kin17; ORFNames=RP11-264C14.1-001;				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;				
OC	Homo				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	NUCLEOTIDE SEQUENCE.				
RC	TISSUE=Testis;				
RX	MEDLINE=20419742; PubMed=10964102; DOI=10.1093/carcin/21.9.1701;				
RA	Kannouche P., Mauffrey P., Pinon-Lataillade G., Mattei M.G.,				
RA	Sarasin A., Daya-Grosjean L., Angulo J.F.;				
RT	"Molecular cloning and characterization of the human KIN17 cDNA encoding a component of the UVC response that is conserved among metazoans."				
RT	Carcinogenesis 21:1701-1710(2000).				
RL	[2]				
RN	NUCLEOTIDE SEQUENCE.				
RC	TISSUE=Pancreas;				
RX	MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;				
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,				
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,				
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,				
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,				
RA	Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,				
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,				
RA	Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,				
RA	Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,				
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,				
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,				
RA	Vallalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,				
RA	Whiting J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,				
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,				
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,				
RA	Butterfield J.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,				
RA	Schmerch A., Schein J.E., Jones S.J.M., Marra M.A.;				
RT	"Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."				
RT	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).				
RL	[3]				
RN	NUCLEOTIDE SEQUENCE.				
RC	TISSUE=Pancreas;				
RA	Strausberg R.;				
RL	Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.				

Q9Y7X9 schizosacch
Q55M94 cryptococcu
Q5K8H7 cryptococcu
Q4WFG0 aspergillus
Q5AQ80 aspergillus
Q54XH1 dictyosteli
Q8PYN9 methanosarc
Q4NX48 anaeromyxob
Q6N18 xenopus lae
Q5GWA7 idiomarina
Q8XVB3 ralatonia s
P77992 thermococcu
Q9UY24 pyrococcus
Q8U438 pyrococcus

ALIGNMENTS


```

DR MGI; MGI:96676; Kin.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR007087; Znf_C2H2.
DR SMART; SM00355; Znf_C2H2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
FT NON_TER 195 AA; 195
SQ SEQUENCE 195 AA; 23518 MW; 7E112E9CA820BFA6 CRC64;

Query Match 45.5%; Score 179; DB 2; Length 195;
Best Local Similarity 100.0%; Pred. No. 7.7e-173;
Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 PKAIANRIKSGLOKLRWYCOMCQKCRDENGFKCHCMSSHORQLLASENPOQFMDYF 68
Db 9 PKAIANRIKSGLOKLRWYCOMCQKCRDENGFKCHCMSSHORQLLASENPOQFMDYF 68

Qy 69 SEFRNDFLELLRRRFGTKRVHNNVYNEVYSHREHIMNATOWETLDTFTKWLGRGLC 128
Db 69 SEFRNDFLELLRRRFGTKRVHNNVYNEVYSHREHIMNATOWETLDTFTKWLGRGLC 128

Qy 129 KVDETPKGWYIQYIDRDPETIRRQLELEKKKKQDLDEEKTAKFIEBQVRRGLEGKEQ 187
Db 129 KVDETPKGWYIQYIDRDPETIRRQLELEKKKKQDLDEEKTAKFIEBQVRRGLEGKEQ 187

RESULT 3
Q8K339 MOUSE PRELIMINARY; PRT; 391 AA.
ID Q8K339 MOUSE PRELIMINARY; PRT; 391 AA.
AC Q8K339;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Antigenic determinant of rec-A protein.
GN Name=Kin;
OS Mus musculus (Mouse);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE!
RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Schnerch A., Schein J.E., Jones S.J.W., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE!
RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RA Strausberg R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE!
RX MEDLINE=92020193; PubMed=1923796;

```

```

RA Angulo J., Rouer E., Mazin A., Mattei M., Tissier A., Horellou P.,
RA Benarous R., Devoret R.;
RT "Identification and expression of the cDNA of KIN17, a zinc-finger
gene located on mouse chromosome 2, encoding a new DNA binding
protein.";
RL Nucleic Acids Res. 19:5117-5123(1991).
DR EMBL; BC028860; AAH28860.1; -; mRNA.
DR PIR; S18666; S18666.
DR Ensembl; ENSMUSG00000037262; Mus musculus.
DR MGI; MGI:96676; Kin.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR005824; KOW.
DR InterPro; IPR006646; KOW sub.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00467; KOW; 1.
DR SMART; SM00739; KOW; 1.
DR PROSITE; PS00355; Znf_C2H2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
KW Nuclear protein.
SQ SEQUENCE 391 AA; 44722 MW; 955BC6A4FF4D3B6E CRC64;

Query Match 45.5%; Score 179; DB 2; Length 391;
Best Local Similarity 100.0%; Pred. No. 1.4e-172;
Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 PKAIANRIKSGLOKLRWYCOMCQKCRDENGFKCHCMSSHORQLLASENPOQFMDYF 68
Db 9 PKAIANRIKSGLOKLRWYCOMCQKCRDENGFKCHCMSSHORQLLASENPOQFMDYF 68

Qy 69 SEFRNDFLELLRRRFGTKRVHNNVYNEVYSHREHIMNATOWETLDTFTKWLGRGLC 128
Db 69 SEFRNDFLELLRRRFGTKRVHNNVYNEVYSHREHIMNATOWETLDTFTKWLGRGLC 128

Qy 129 KVDETPKGWYIQYIDRDPETIRRQLELEKKKKQDLDEEKTAKFIEBQVRRGLEGKEQ 187
Db 129 KVDETPKGWYIQYIDRDPETIRRQLELEKKKKQDLDEEKTAKFIEBQVRRGLEGKEQ 187

RESULT 4
Q6GL97_XENTR PRELIMINARY; PRT; 389 AA.
ID Q6GL97_XENTR PRELIMINARY; PRT; 389 AA.
AC Q6GL97;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE MGC69184 protein.
GN Name=MGC69184;
OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus; Silurana.
OX NCBI_TaxID=8364;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Embryo;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

```

```

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Embryo;
RA Klein S., Gerhard D.S.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC074606; AAH74606.1; -; mRNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR007087; Znf_C2H2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN 1.
SQ SEQUENCE 389 AA; 44987 MW; 793E70C997451FEA CRC64;

Query Match 16.5%; Score 65; DB 2; Length 389;
Best Local Similarity 100.0%; Pred. No. 1e-56;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DFLTPKAIRNRIKSGLOKLRWYQMCQKQCRDENGFKCHMSHSHORQLLASNPQOF 64
DB 5 DFLTPKAIRNRIKSGLOKLRWYQMCQKQCRDENGFKCHMSHSHORQLLASNPQOF 64

QY 65 MDYFS 69
DB 65 MDYFS 69

RESULT 5
Q68F56_XENLA
ID Q68F56_XENLA PRELIMINARY; PRT; 387 AA.
AC Q68F56;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE MGC81626 protein.
GN Name=MGC81626;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae;
OC Xenopodinae; Xenopus; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Embryo;
RX MEDLINE=223341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative.";
RL Dev. Dyn. 225:384-391(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Embryo;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Heiton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human

```

```

RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Embryo;
RA Klein S., Gerhard D.S.;
RL Submitted (AUG-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC079988; AAH79988.1; -; mRNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR005824; KOW.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00467; KOW; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
SQ SEQUENCE 387 AA; 44491 MW; 4CC8B56B047F3779 CRC64;

Query Match 12.7%; Score 50; DB 2; Length 387;
Best Local Similarity 100.0%; Pred. No. 1.8e-41;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 NRIKSGLOKLRWYQMCQKQCRDENGFKCHMSHSHORQLLASNPQ 63
DB 14 NRIKSGLOKLRWYQMCQKQCRDENGFKCHMSHSHORQLLASNPQ 63

RESULT 6
Q4S5G2_TETNG
ID Q4S5G2_TETNG PRELIMINARY; PRT; 427 AA.
AC Q4S5G2;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Chromosome 19 SCAP14731, whole genome shotgun sequence.
GN ORFNames=GSTENG00023750001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Jallion O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segreus B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castellani V., Katinka M., Vacherie B.,
RA Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
RA Parra G., Lardier G., Chapelle C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier P., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CAAE0104731; CAG04120.1; -; Genomic DNA.
SQ SEQUENCE 427 AA; 48007 MW; ADF91299BEC23B CRC64;

Query Match 12.2%; Score 48; DB 2; Length 427;
Best Local Similarity 100.0%; Pred. No. 2.1e-39;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```
QY 14 NRKSKGLQKLRWYCMCKQCRDENGFKCHMSHQRQLLASNP 61
DB 14 NRKSKGLQKLRWYCMCKQCRDENGFKCHMSHQRQLLASNP 61

RESULT 7
Q5RH25_BRARE PRELIMINARY; PRT; 383 AA.
AC Q5RH25;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Novel protein similar to vertebrate KIN, antigenic determinant of reca
DE protein homolog (Mouse) (KIN).
GN Name:OTTDPARP0000008158; ORFNames=CH211-235A22.3-001;
OS Brachydanio rerio (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN 1;
RP NUCLEOTIDE SEQUENCE.
RA Hunter G.;
RL Submitted (DEC-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BX640594; CAIL1855.1; -; Genomic_DNA.
DR InterPro; IPR005824; KOW.
DR Pfam; PF00467; KOW; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKOWN 1.
SQ SEQUENCE 383 AA; 44337 MW; 1F0AA34FB62176D1 CRC64;

Query Match 10.9%; Score 43; DB 2; Length 383;
Best Local Similarity 100.0%; Pred. No. 2.3e-34;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRWYCMCKQCRDENGFKCHMSHQRQLLASNP 61
DB 19 KGLQKLRWYCMCKQCRDENGFKCHMSHQRQLLASNP 61

RESULT 8
Q4TDV9_TETNG PRELIMINARY; PRT; 269 AA.
AC Q4TDV9;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Chromosome undetermined SCAF6005, whole genome shotgun sequence.
DE (Fragment).
GN ORFNames=GSTENG0002601001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN 1;
RP NUCLEOTIDE SEQUENCE.
RA Jallou O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Bossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anhouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA Bionet C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Craud C., Duprat S., Brottier P., Coutanceau J.P., Guzy J.,
RA Parra G., Lardier G., Chaplier C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissenbach J., Roest Crolius H.,
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
the early vertebrate proto-karyotype."
```

```
RL Nature 431:946-957 (2004).
RN 1;
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CAAE01006005; CAF88923.1; -; Genomic_DNA.
FT NON_TER 1 1
FT NON_TER 269 269
SQ SEQUENCE 269 AA; 31736 MW; E6883839A2CD17AF CRC64;

Query Match 10.4%; Score 41; DB 2; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.8e-32;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 WETLTFTKWLGRGKCKVDETPKGVYQIYIDRDPETIRQ 152
DB 112 WETLTFTKWLGRGKCKVDETPKGVYQIYIDRDPETIRQ 98

RESULT 9
Q9AY87_ORYSA PRELIMINARY; PRT; 402 AA.
AC Q9AY87;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein OSUNBa0004B24.17.
GN Name=OSUNBa0004B24.17;
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN 1;
RP NUCLEOTIDE SEQUENCE.
RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Burr P.C., Hsiao J.,
RA Zisman V., Pal G., Bowman C.L., Fujii C.Y., Vanaken S.E.,
RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblyum T.V.,
RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN 2;
RP NUCLEOTIDE SEQUENCE.
RA Buell R.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC084319; AAG59654.1; -; Genomic_DNA.
DR Gramene; Q9AY87; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR007087; Znf_C2H2.
DR SMART; SM00355; Znf_C2H2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKOWN 1.
KW Hypothetical protein.
SQ SEQUENCE 402 AA; 45962 MW; A1F4E38ECC60D13E CRC64;

Query Match 8.9%; Score 35; DB 2; Length 402;
Best Local Similarity 100.0%; Pred. No. 3.2e-26;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRWYCMCKQCRDENGFKCHMSHQRQ 53
DB 19 KGLQKLRWYCMCKQCRDENGFKCHMSHQRQ 53

RESULT 10
Q9ZVU5_ARATH PRELIMINARY; PRT; 411 AA.
ID Q9ZVU5_ARATH PRELIMINARY;
AC Q9ZVU5;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
```

DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)
 DT 01-FEB-2005 (T-EMBLrel. 29, Last annotation update)
 DE Names=TS14.13; Synonyms=At1G55460;
 GN Arabidopsis thaliana (Mouse-ear cress).
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopses.
 OX NCBI_TaxID=3702;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
 RA Altafi H., Araujo R., Huizar L., Rowley D., Buehler E., Dunn P.,
 RA Gonzalez A., Kremenetskaia I., Kim C., Lenz C., Li J., Liu S.,
 RA Luros S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,
 RA Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;
 RN Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RP NUCLEOTIDE SEQUENCE.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Egu P., Lee J.M.,
 RA Toriumi M., Yu G., Brooks S., Chao Q., Chen H., Karlin-Neumann G.,
 RA Kim C., Lam B., Miranda M., Nguyen M., Palm C.J., Shinn P.,
 RA Southwick A., Davis R.W., Ecker J.R., Theologis A.;
 RN Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 RL [3]
 RP NUCLEOTIDE SEQUENCE.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 RA Goldsmith A.D., Lee J.M., Quach H.L., Tortumi M., Yu G., Bowser L.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
 RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RN Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AC005223; AAD10649.1; -; Genomic_DNA.
 DR EMBL; AY051011; AAK93688.1; -; mRNA.
 DR EMBL; AF360132; AAK25842.1; -; mRNA.
 DR PIR; H95596; H95596.
 DR GO; GO:0005634; C:nucleic acid binding; IEA.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR InterPro; IPR005824; KOW.
 DR InterPro; IPR006646; KOW-sub.
 DR PIR; PF00467; KOW; 1.
 DR SMART; SM00739; KOW; 1.
 DR SMART; SM00355; Znf_C2H2; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
 KW Hypothetical protein.
 SQ SEQUENCE 411 AA; 47288 MW; 9DA6F8648002065D CRC64;
 Query Match 8.9%; Score 35; DB 2; Length 411;
 Best Local Similarity 100.0%; Pred. No. 3.3e-26;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHQ 53
 DB 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHQ 53
 RESULT 11
 Q75LU5_ORYSA
 ID Q75LU5_ORYSA PRELIMINARY; PRT; 430 AA.
 AC Q75LU5;
 DT 05-JUL-2004 (T-EMBLrel. 27, Created)
 DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)
 DE Hypothetical protein OSUNB0015121.3.
 GN Names=OSUNB0015121.3;
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.

OX NCBI_TaxID=39947;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,
 RA Overton II L.L., Tsitrin T., Kim M.M., Bera J.J., Jin S.S.,
 RA Fadros D.W., Tallon L.J., Koo H., Ziamann V., Hsiao J., Blunt S.,
 RA Vanaken S.S., Riedmuller S.B., Utterback T.T., Feidblyum T.V.,
 RA Yang Q.Q., Haas B.J., Suh B.B., Peterson J.J., Quackenbush J.,
 RA White O., Salzberg S.L., Fraser C.W.;
 RN Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RP NUCLEOTIDE SEQUENCE.
 RA Buell R.;
 RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC091302; AAR00634.1; -; Genomic_DNA.
 DR Gramene; Q75LU5; -.
 DR InterPro; IPR005824; KOW.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00467; KOW; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
 KW Hypothetical protein.
 SQ SEQUENCE 430 AA; 49067 MW; 6151D748ED2F4690 CRC64;
 Query Match 8.9%; Score 35; DB 2; Length 430;
 Best Local Similarity 100.0%; Pred. No. 3.4e-26;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHQ 53
 DB 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHQ 53
 RESULT 12
 Q4TDW0_TETNG
 ID Q4TDW0_TETNG PRELIMINARY; PRT; 88 AA.
 AC Q4TDW0;
 DT 13-SEP-2005 (T-EMBLrel. 31, Created)
 DT 13-SEP-2005 (T-EMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (T-EMBLrel. 31, Last annotation update)
 DE Chromosome undetermined SCAF6004, whole genome shotgun sequence.
 DE (Fragment).
 DE ORFNames=GSTENG0002600001;
 OS Tetraodon nigroviridis (Green puffer).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 OC Tetraodontidae; Tetraodontidae; Tetraodon.
 OX NCBI_TaxID=99883;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
 RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
 RA Nicaud S., Jaffe D., Fiecher S., Lutfalla G., Dossat C., Segurens B.,
 RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
 RA Anthouard V., Jubin C., Caetelli V., Katinka M., Vacherie B.,
 RA Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
 RA Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
 RA Parra G., Lardier G., Chapelle C., McKernan K.J., McEwan P., Bosak S.,
 RA Kellis M., Volff J.N., Guigio R., Zody M.C., Mesirov J.,
 RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
 RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
 RA Wincker P., Lander E.S., Weissbach J., Roest Crolius H.;
 RA "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
 RT the early vertebrate proto-karyotype.";
 RN Nature 431:946-957(2004).
 RL [2]
 RP NUCLEOTIDE SEQUENCE.
 RG Genoscope; Whitehead Institute Centre for Genome Research;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 CC -1- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL; CAAB01006004; CAF88922.1; -; Genomic_DNA.

```
FT NON_TER 1
FT NON_TER 88
SQ SEQUENCE 88 AA; 9709 MW; C5BD169AF081CBC2 CRC64;

Query Match 7.4%; Score 29; DB 2; Length 88;
Best Local Similarity 100.0%; Pred.No. 1e-20;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 314 AVVMIDSGDKLKDQTHLETVIPAPGKR 342
Db 10 AVVMIDSGDKLKDQTHLETVIPAPGKR 38

RESULT 13
Q6XI90_DROYA
ID Q6XI90_DROYA PRELIMINARY; PRT; 178 AA.
AC Q6XI90;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Similar to Drosophila melanogaster Kin17 (Fragment).
OS Drosophila yakuba (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7245;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22887302; PubMed=14525923; DOI=10.1101/gr.1311003;
RA Domazet-Looso T.; Tautz D.;
RT "An evolutionary analysis of orphan genes in Drosophila.";
RL Genome Res. 13:2213-2219(2003).
DR EMBL; AY231941; AAR09964.1; -; mRNA.
DR InterPro; IPR007087; Znf_C2H2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 178
SQ SEQUENCE 178 AA; 21188 MW; B251B8630CB76C3 CRC64;

Query Match 3.8%; Score 15; DB 2; Length 178;
Best Local Similarity 100.0%; Pred.No. 3.2e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 KSKGLQKLRWYQCMC 31
Db 17 KSKGLQKLRWYQCMC 31

RESULT 14
O76926_DROME
ID O76926_DROME PRELIMINARY; PRT; 244 AA.
AC O76926;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE KIN17 protein (Fragment).
GN Name=kin17; ORFNames=CG5649;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=EST28;
RA Garcia V., de La Roche Saint Andre C.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AJ006528; CAA07089.1; -; Genomic DNA.
DR Ensembl; CG5649; Drosophila melanogaster.
DR FlyBase; FBgn0024887; CG5649.
DR FlyBase; FBgn0024887; kin17.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
```

This Page Blank (uspic)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

QM protein - protein search, using sw model

Run on: November 23, 2005, 16:35:49 ; Search time 33.0842 Seconds
(without alignments)
982.087 Million cell updates/sec

Title: us-09-555-529-26

Perfect score: 393

Sequence: 1 MGKSDFLTPKAIANRIKSG.....GPKGRVRGVIQYEDISKLA 393

Scoring table:

Gapop 60;0 , Gapext 60.0

Searched: 572060 seqs, 82675679 residues

Word size : 0

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:..
1: /cgn2_6/ptodata/1/iaa/5 COMB.pdp.*
2: /cgn2_6/ptodata/1/iaa/6 COMB.pdp.*
3: /cgn2_6/ptodata/1/iaa/H COMB.pdp.*
4: /cgn2_6/ptodata/1/iaa/PTCUS COMB.pdp.*
5: /cgn2_6/ptodata/1/iaa/RE COMB.pdp.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pdp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	2.3	401	2	US-09-270-767-44680
2	8	2.0	55	2	US-09-513-999C-6360
3	8	2.0	61	2	US-09-149-476-615
4	8	2.0	136	2	US-09-270-767-32008
5	8	2.0	176	1	US-08-809-267-10
6	8	2.0	176	2	US-10-135-807-8
7	8	2.0	176	4	PCT-US95-13662A-10
8	8	2.0	178	2	US-10-135-807-6
9	8	2.0	178	2	US-10-135-807-7
10	8	2.0	984	2	US-09-171-461-14
11	8	2.0	984	2	US-09-970-711-14
12	7	1.8	61	2	US-09-248-796A-22505
13	7	1.8	63	2	US-08-971-089-2
14	7	1.8	63	2	US-09-513-999C-5320
15	7	1.8	63	2	US-10-117-604A-2
16	7	1.8	64	2	US-09-248-796A-26971
17	7	1.8	67	2	US-09-248-796A-24256
18	7	1.8	68	2	US-09-270-767-34723
19	7	1.8	68	2	US-09-270-767-49940
20	7	1.8	69	2	US-09-107-433-4761
21	7	1.8	76	2	US-09-248-796A-21942
22	7	1.8	77	2	US-09-248-796A-27087
23	7	1.8	87	2	US-09-248-796A-25044
24	7	1.8	90	2	US-09-613-486-24
25	7	1.8	91	2	US-09-902-540-12006
26	7	1.8	94	2	US-09-621-976-5665
27	7	1.8	112	2	US-09-543-681A-8319

ALIGNMENTS

RESULT 1

US-09-270-767-44680
; Sequence 44680, Application US/09270767
; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 44680

; LENGTH: 401

; TYPE: PRT

; ORGANISM: *Drosophila melanogaster*

; FEATURE:

; OTHER INFORMATION: Xaa means any amino acid

US-09-270-767-44680

Query Match 2.3%; Score 9; DB 2; Length 401;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 244 SSQSSTQSK 252

Db 113 SSQSSTQSK 121

RESULT 2

US-09-513-999C-6360
; Sequence 6360, Application US/09513999C
; Patent No. 6783961

; GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.

; APPLICANT: Duclert, A.

; APPLICANT: Giordano, J.Y.

; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.

; Patent No. 6783961

; FILE REFERENCE: 59 US2 REG

; CURRENT APPLICATION NUMBER: US/09/513,999C

; CURRENT FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/122,487

; PRIOR FILING DATE: 1999-02-26

; NUMBER OF SEQ ID NOS: 36681

; SOFTWARE: Patent.Pm

; SEQ ID NO 6360

; LENGTH: 55

; TYPE: PRT

; ORGANISM: *Homo sapiens*

Sequence 20, Appl
Sequence 28432, A
Sequence 44294, A
Sequence 15608, A
Sequence 8310, Ap
Sequence 8236, Ap
Sequence 8301, Ap
Sequence 41907, A
Sequence 2, Appl
Sequence 16, Appl
Sequence 26, Appl
Sequence 30185, A
Sequence 15986, A
Sequence 6928, Ap
Sequence 14, Appl
Sequence 14, Appl
Sequence 46506, A
Sequence 8425, Ap

28 7 1.8 120 2 US-09-311-021-20
29 7 1.8 126 2 US-09-252-991A-28432
30 7 1.8 143 2 US-09-270-767-44294
31 7 1.8 152 2 US-09-248-796A-15608
32 7 1.8 160 2 US-09-543-681A-8310
33 7 1.8 178 2 US-09-489-039A-8236
34 7 1.8 182 2 US-09-489-039A-8301
35 7 1.8 212 2 US-09-270-767-41907
36 7 1.8 214 2 US-09-391-741A-2
37 7 1.8 214 2 US-09-391-741A-16
38 7 1.8 214 2 US-09-391-741A-26
39 7 1.8 249 2 US-09-252-991A-30185
40 7 1.8 299 2 US-09-248-796A-15986
41 7 1.8 347 2 US-09-949-016-6928
42 7 1.8 352 2 US-09-073-009-14
43 7 1.8 352 2 US-09-073-010-14
44 7 1.8 365 2 US-09-270-767-46506
45 7 1.8 371 2 US-09-949-016-8425

```
/ ; FEATURE:
/ ; NAME/KEY: UNSURE
/ ; LOCATION: 8
/ ; OTHER INFORMATION: Xaa=Leu or Pro
/ ; FEATURE:
/ ; NAME/KEY: UNSURE
/ ; LOCATION: 11
/ ; OTHER INFORMATION: Xaa=Glu or Lys
US-09-513-999C-6360

Query Match          2.0%; Score 8; DB 2; Length 55;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 251 SKEKKKK 258
Db 29 SKEKKKK 36

RESULT 3
US-09-149-476-615
/ ; Sequence 615, Application US/09149476
/ ; Patent No. 6420526
/ ; GENERAL INFORMATION:
/ ; APPLICANT: Rosen et al.
/ ; TITLE OF INVENTION: 186 Human Secreted proteins
/ ; FILE REFERENCE: PZ002P1
/ ; CURRENT APPLICATION NUMBER: US/09/149,476
/ ; CURRENT FILING DATE: 1998-09-08
/ ; EARLIER APPLICATION NUMBER: PCT/US98/04493
/ ; EARLIER FILING DATE: 1998-03-06
/ ; EARLIER APPLICATION NUMBER: 60/040,162
/ ; EARLIER FILING DATE: 1997-03-07
/ ; EARLIER APPLICATION NUMBER: 60/040,333
/ ; EARLIER FILING DATE: 1997-03-07
/ ; EARLIER APPLICATION NUMBER: 60/038,621
/ ; EARLIER FILING DATE: 1997-03-07
/ ; EARLIER APPLICATION NUMBER: 60/040,626
/ ; EARLIER FILING DATE: 1997-03-07
/ ; EARLIER APPLICATION NUMBER: 60/040,334
/ ; EARLIER FILING DATE: 1997-03-07
/ ; EARLIER APPLICATION NUMBER: 60/040,336
/ ; EARLIER FILING DATE: 1997-03-07
/ ; EARLIER APPLICATION NUMBER: 60/040,163
/ ; EARLIER FILING DATE: 1997-03-07
/ ; EARLIER APPLICATION NUMBER: 60/047,600
/ ; EARLIER FILING DATE: 1997-05-23
/ ; EARLIER APPLICATION NUMBER: 60/047,615
/ ; EARLIER FILING DATE: 1997-05-23
/ ; EARLIER APPLICATION NUMBER: 60/047,597
/ ; EARLIER FILING DATE: 1997-05-23
/ ; EARLIER APPLICATION NUMBER: 60/047,502
/ ; EARLIER FILING DATE: 1997-05-23
/ ; EARLIER APPLICATION NUMBER: 60/047,633
/ ; EARLIER FILING DATE: 1997-05-23
/ ; EARLIER APPLICATION NUMBER: 60/047,583
/ ; EARLIER FILING DATE: 1997-05-23
/ ; EARLIER APPLICATION NUMBER: 60/047,617
/ ; EARLIER FILING DATE: 1997-05-23
/ ; EARLIER APPLICATION NUMBER: 60/047,618
/ ; EARLIER FILING DATE: 1997-05-23
/ ; EARLIER APPLICATION NUMBER: 60/047,503
/ ; EARLIER FILING DATE: 1997-05-23
/ ; EARLIER APPLICATION NUMBER: 60/047,592
/ ; EARLIER FILING DATE: 1997-05-23
/ ; EARLIER APPLICATION NUMBER: 60/047,581
/ ; EARLIER FILING DATE: 1997-05-23
/ ; EARLIER APPLICATION NUMBER: 60/047,584
/ ; EARLIER FILING DATE: 1997-05-23
/ ; EARLIER APPLICATION NUMBER: 60/047,500
/ ; EARLIER FILING DATE: 1997-05-23
/ ; EARLIER APPLICATION NUMBER: 60/047,587
/ ; EARLIER FILING DATE: 1997-05-23
/ ; EARLIER APPLICATION NUMBER: 60/047,587
/ ; EARLIER FILING DATE: 1997-05-23
/ ; EARLIER APPLICATION NUMBER: 60/047,492
/ ; EARLIER FILING DATE: 1997-05-23
/ ; EARLIER APPLICATION NUMBER: 60/047,598
/ ; EARLIER FILING DATE: 1997-05-23
/ ; EARLIER APPLICATION NUMBER: 60/047,613
/ ; EARLIER FILING DATE: 1997-05-23
/ ; EARLIER APPLICATION NUMBER: 60/047,582
/ ; EARLIER FILING DATE: 1997-05-23
/ ; EARLIER APPLICATION NUMBER: 60/047,596
/ ; EARLIER FILING DATE: 1997-05-23
/ ; EARLIER APPLICATION NUMBER: 60/047,612
/ ; EARLIER FILING DATE: 1997-05-23
/ ; EARLIER APPLICATION NUMBER: 60/047,632
/ ; EARLIER FILING DATE: 1997-05-23
/ ; EARLIER APPLICATION NUMBER: 60/047,601
/ ; EARLIER FILING DATE: 1997-05-23
/ ; EARLIER APPLICATION NUMBER: 60/043,580
/ ; EARLIER FILING DATE: 1997-04-11
/ ; EARLIER APPLICATION NUMBER: 60/043,568
/ ; EARLIER FILING DATE: 1997-04-11
/ ; EARLIER APPLICATION NUMBER: 60/043,314
/ ; EARLIER FILING DATE: 1997-04-11
/ ; EARLIER APPLICATION NUMBER: 60/043,569
/ ; EARLIER FILING DATE: 1997-04-11
/ ; EARLIER APPLICATION NUMBER: 60/043,311
/ ; EARLIER FILING DATE: 1997-04-11
/ ; EARLIER APPLICATION NUMBER: 60/043,671
/ ; EARLIER FILING DATE: 1997-04-11
/ ; EARLIER APPLICATION NUMBER: 60/043,674
/ ; EARLIER FILING DATE: 1997-04-11
/ ; EARLIER APPLICATION NUMBER: 60/043,669
/ ; EARLIER FILING DATE: 1997-04-11
/ ; EARLIER APPLICATION NUMBER: 60/043,312
/ ; EARLIER FILING DATE: 1997-04-11
/ ; EARLIER APPLICATION NUMBER: 60/043,313
/ ; EARLIER FILING DATE: 1997-04-11
/ ; EARLIER APPLICATION NUMBER: 60/043,672
/ ; EARLIER FILING DATE: 1997-04-11
/ ; EARLIER APPLICATION NUMBER: 60/043,315
/ ; EARLIER FILING DATE: 1997-04-11
/ ; EARLIER APPLICATION NUMBER: 60/048,974
/ ; EARLIER FILING DATE: 1997-06-06
/ ; EARLIER APPLICATION NUMBER: 60/056,886
/ ; EARLIER FILING DATE: 1997-08-22
/ ; EARLIER APPLICATION NUMBER: 60/056,877
/ ; EARLIER FILING DATE: 1997-08-22
/ ; EARLIER APPLICATION NUMBER: 60/056,889
/ ; EARLIER FILING DATE: 1997-08-22
/ ; EARLIER APPLICATION NUMBER: 60/056,893
/ ; EARLIER FILING DATE: 1997-08-22
/ ; EARLIER APPLICATION NUMBER: 60/056,630
/ ; EARLIER FILING DATE: 1997-08-22
/ ; EARLIER APPLICATION NUMBER: 60/056,878
/ ; EARLIER FILING DATE: 1997-08-22
/ ; EARLIER APPLICATION NUMBER: 60/056,662
/ ; EARLIER FILING DATE: 1997-08-22
/ ; EARLIER APPLICATION NUMBER: 60/056,872
/ ; EARLIER FILING DATE: 1997-08-22
/ ; EARLIER APPLICATION NUMBER: 60/056,882
/ ; EARLIER FILING DATE: 1997-08-22
/ ; EARLIER APPLICATION NUMBER: 60/056,637
/ ; EARLIER FILING DATE: 1997-08-22
/ ; EARLIER APPLICATION NUMBER: 60/056,903
/ ; EARLIER FILING DATE: 1997-08-22
/ ; EARLIER APPLICATION NUMBER: 60/056,888
/ ; EARLIER FILING DATE: 1997-08-22
/ ; EARLIER APPLICATION NUMBER: 60/056,879
/ ; EARLIER FILING DATE: 1997-08-22
/ ; EARLIER APPLICATION NUMBER: 60/056,880
/ ; EARLIER FILING DATE: 1997-08-22
/ ; EARLIER APPLICATION NUMBER: 60/056,894
/ ; EARLIER FILING DATE: 1997-08-22
/ ; EARLIER APPLICATION NUMBER: 60/056,911
```


; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,636
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,874
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,910
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,864
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,631
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,845
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,892
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/057,761
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/047,595
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,599
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,588
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,585
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,586
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,590
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,594
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,589
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,593
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,614
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,578
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,576
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/047,501
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,670
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/056,632
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,664
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,876
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,881
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,909
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,875
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,862
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,887
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,908
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/048,964
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/057,650
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/056,884
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/057,669
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/049,610
; EARLIER FILING DATE: 1997-06-13

; EARLIER APPLICATION NUMBER: 60/061,060
; EARLIER FILING DATE: 1997-10-02

Query Match 2.0%; Score 8; DB 2; Length 61;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 244 SSSSTQS 251
Db 44 SSSSTQS 51
|||||||

RESULT 4
US-09-270-767-32008
; Sequence 32008, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 64517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32008
; LENGTH: 136
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-32008

Query Match 2.0%; Score 8; DB 2; Length 136;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 331 HLETVIPA 338
Db 72 HLETVIPA 79
|||||||

RESULT 5
US-08-809-267-10
; Sequence 10, Application US/0809267
; Patent No. 5861296
; GENERAL INFORMATION:
; APPLICANT: LENNOX, Tricia L.
; APPLICANT: SLATKO, Barton E.
; APPLICANT: SEARS, Lauren E.
; TITLE OF INVENTION: PURIFIED THERMOSTABLE INORGANIC
; TITLE OF INVENTION: PYROPHOSPHATASE OBTAINABLE FROM THERMOCOCCUS
; TITLE OF INVENTION: LITORALIS
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEW ENGLAND BIOLABS, INC.
; STREET: 32 TOZER ROAD
; CITY: BEVERLY
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 01915
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/809,267
; FILING DATE: 12-MAR-1997
; CLASSIFICATION: 433
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13662
; FILING DATE:
; APPLICATION NUMBER: US 08/329,721
; FILING DATE: 25-OCT-1994
; ATTORNEY/AGENT INFORMATION:

```
; NAME: WILLIAMS, GREGORY D.
; REGISTRATION NUMBER: 30901
; REFERENCE/DOCKET NUMBER: NEB-105-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 927-5054
; TELEFAX: (508) 927-1705
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 176 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-809-267-10

Query Match      2.0%; Score 8; DB 1; Length 176;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      317 KMIDSGDK 324
Db      95 KMIDSGDK 102

RESULT 6
US-10-135-807-8
; Sequence 8, Application US/10135807
; Patent No. 6951744
; GENERAL INFORMATION:
; APPLICANT: The Secretary of State for Defence in Her Britannic Majesty's
; APPLICANT: Government of the United Kingdom of Great Britain and
; APPLICANT: No. 6951744thern Ireland
; APPLICANT: Clark, Duncan Roy
; APPLICANT: Vincent, Suzanne P
; TITLE OF INVENTION: Amplification process
; FILE REFERENCE: CG/P/133/WOD
; CURRENT APPLICATION NUMBER: US/10/135,807
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: GB 0110501.4
; PRIOR FILING DATE: 2001-04-30
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Thermococcus litoralis
US-10-135-807-8

Query Match      2.0%; Score 8; DB 2; Length 176;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      317 KMIDSGDK 324
Db      95 KMIDSGDK 102

RESULT 7
PCT-US95-13662A-10
; Sequence 10, Application PC/TUS9513662A
; GENERAL INFORMATION:
; APPLICANT: LENNOX, Tricia L.
; APPLICANT: SLATKO, Barton E.
; APPLICANT: SEARS, Lauren E.
; TITLE OF INVENTION: PURIFIED THERMOSTABLE INORGANIC
; TITLE OF INVENTION: PYROPHOSPHATASE OBTAINABLE FROM THERMOCOCCUS
; TITLE OF INVENTION: LITORALIS
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEW ENGLAND BIOLABS, INC.
; STREET: 32 TOZER ROAD
; CITY: BEVERLY
; STATE: MASSACHUSETTS
; COUNTRY: USA
```

```
; ZIP: 01915
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13662A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,721
; FILING DATE: 25-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAMS, GREGORY D.
; REGISTRATION NUMBER: 30901
; REFERENCE/DOCKET NUMBER: NEB-105-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 927-5054
; TELEFAX: (508) 927-1705
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 176 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-13662A-10

Query Match      2.0%; Score 8; DB 4; Length 176;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      317 KMIDSGDK 324
Db      95 KMIDSGDK 102

RESULT 8
US-10-135-807-6
; Sequence 6, Application US/10135807
; Patent No. 6951744
; GENERAL INFORMATION:
; APPLICANT: The Secretary of State for Defence in Her Britannic Majesty's
; APPLICANT: Government of the United Kingdom of Great Britain and
; APPLICANT: No. 6951744thern Ireland
; APPLICANT: Clark, Duncan Roy
; APPLICANT: Vincent, Suzanne P
; TITLE OF INVENTION: Amplification process
; FILE REFERENCE: CG/P/133/WOD
; CURRENT APPLICATION NUMBER: US/10/135,807
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: GB 0110501.4
; PRIOR FILING DATE: 2001-04-30
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-10-135-807-6

Query Match      2.0%; Score 8; DB 2; Length 178;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      317 KMIDSGDK 324
Db      95 KMIDSGDK 102

RESULT 9
US-10-135-807-7
; Sequence 7, Application US/10135807
```

; Patent No. 6951744
; GENERAL INFORMATION:
; APPLICANT: The Secretary of State for Defence in Her Britannic Majesty's
; APPLICANT: Government of the United Kingdom of Great Britain and
; APPLICANT: No. 6951744thern Ireland
; APPLICANT: Clark, Duncan Roy
; APPLICANT: Vincent, Suzanne P
; TITLE OF INVENTION: Amplification process
; FILE REFERENCE: CG/P/133/WOD
; CURRENT APPLICATION NUMBER: US/10/135,807
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: GB 0110501.4
; PRIOR FILING DATE: 2001-04-30
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Pyrococcus abyssi
US-10-135-807-7

Query Match 2.0%; Score 8; DB 2; Length 178;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 317 KMIDSGDK 324
Db 95 KMIDSGDK 102

RESULT 10
US-09-171-461-14
; Sequence 14, Application US/09171461
; Patent No. 6335016
; GENERAL INFORMATION:
; APPLICANT: Baker, Adam
; APPLICANT: Cotten, Matthew
; APPLICANT: Chiocca, Susanna
; APPLICANT: Kurzbauer, Robert
; TITLE OF INVENTION: Chicken Embryo Lethal Orphan (CELO) Virus
; FILE REFERENCE: 0652.1800000
; CURRENT APPLICATION NUMBER: US/09/171,461
; CURRENT FILING DATE: 1999-01-12
; EARLIER APPLICATION NUMBER: PCT/EP97/01944
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 984
; TYPE: PRT
; ORGANISM: CELO Virus
; FEATURE:
; OTHER INFORMATION: Position: 23680..26634 /gene: L4 /product: L4 100K
US-09-171-461-14

Query Match 2.0%; Score 8; DB 2; Length 984;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 FLTPKAIA 13
Db 301 FLTPKAIA 308

RESULT 11
US-09-970-711-14
; Sequence 14, Application US/09970711
; Patent No. 6773709
; GENERAL INFORMATION:
; APPLICANT: Baker, Adam
; APPLICANT: Cotten, Matthew
; APPLICANT: Chiocca, Susanna

; APPLICANT: Kurzbauer, Robert
; APPLICANT: Schaffner, Gotthold
; TITLE OF INVENTION: Chicken Embryo Lethal Orphan (CELO) Virus
; FILE REFERENCE: 0652.1800001
; CURRENT APPLICATION NUMBER: US/09/970,711
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/171,461
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: PCT/EP97/01944
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 984
; TYPE: PRT
; ORGANISM: CELO Virus
; FEATURE:
; OTHER INFORMATION: Position: 23680..26634 /gene: L4 /product: L4 100K
US-09-970-711-14

Query Match 2.0%; Score 8; DB 2; Length 984;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 FLTPKAIA 13
Db 301 FLTPKAIA 308

RESULT 12
US-09-248-796A-22505
; Sequence 22505, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 22505
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-22505

Query Match 1.8%; Score 7; DB 2; Length 61;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 252 KEKKKKK 258
Db 4 KEKKKKK 10

RESULT 13
US-08-971-089-2
; Sequence 2, Application US/08971089
; Patent No. 6376174
; GENERAL INFORMATION:
; APPLICANT: Pulat, Stefan M.
; APPLICANT: Scoles, Daniel R.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING
; TITLE OF INVENTION: SCHWANNOMIN-BINDING-PROTEINS AND PRODUCTS RELATED THEREO
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL & FLORES, LLP
; STREET: 4370 La Jolla Village Drive, Suite 700

```
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/971,089
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/030,987
; FILING DATE: 15-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ramos, Robert T.
; REGISTRATION NUMBER: 37,915
; REFERENCE/DOCKET NUMBER: P-CE 2862
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)535-9001
; TELEFAX: (619)535-8949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 63 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-971-089-2

Query Match 1.8%; Score 7; DB 2; Length 63;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 KEKXXXX 258
DB 14 KEKXXXX 20
```

```
RESULT 14
US-09-513-999C-5320
; Sequence 5320 Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 5320
; LENGTH: 63
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-513-999C-5320
```

```
Query Match 1.8%; Score 7; DB 2; Length 63;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 KEKXXXX 258
DB 46 KEKXXXX 52
```

. RESULT 15

```
US-10-117-604A-2
; Sequence 2, Application US/10117604A
; Patent No. 6960650
; GENERAL INFORMATION:
; APPLICANT: Pulst, Stefan M.
; APPLICANT: Scoles, Daniel R.
; TITLE OF INVENTION: Nucleic Acid Encoding
; TITLE OF INVENTION: Schwammomin-Binding-Proteins and Products Related Thereto
; FILE REFERENCE: 66783-101
; CURRENT APPLICATION NUMBER: US/10/117,604A
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/030,987
; PRIOR FILING DATE: 1996-11-15
; PRIOR APPLICATION NUMBER: US 08/971,089
; PRIOR FILING DATE: 1997-11-14
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 63
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-117-604A-2

Query Match 1.8%; Score 7; DB 2; Length 63;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 KEKXXXX 258
DB 14 KEKXXXX 20
```

Search completed: November 23, 2005, 16:50:49
Job time : 34.0842 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:44:35 ; Search time 111.283 Seconds
(without alignments)
1475.577 Million cell updates/sec

Title: US-09-555-529-26
Perfect score: 393
Sequence: 1 MGKSDFLTPKAIANRIKSG.....GPLKGRVGEIQYEDISKLA 393

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1867569 seqs, 417829326 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA.Main.*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	393	100.0	393	4	US-10-408-765A-1465
2	73	18.6	84	4	US-10-106-698-5464
3	35	8.9	120	4	US-10-767-701-57151
4	35	8.9	138	4	US-10-437-963-191916
5	35	8.9	423	5	US-10-739-930-8499
6	35	8.9	424	4	US-10-425-115-341788
7	35	8.9	437	4	US-10-425-114-48183
8	35	8.9	437	4	US-10-425-114-58982
9	30	7.6	176	4	US-10-424-599-168326
10	15	3.8	390	6	US-10-097-143-41040
11	9	2.3	146	4	US-10-424-599-160516
12	9	2.3	374	4	US-10-424-599-264493
13	8	2.0	8	4	US-10-022-066-38
14	8	2.0	8	4	US-10-022-066-610
15	8	2.0	39	4	US-10-425-115-357220
16	8	2.0	42	4	US-10-424-599-247175
17	8	2.0	61	3	US-09-809-391-615
18	8	2.0	61	3	US-09-882-171-615
19	8	2.0	61	4	US-10-164-861-615
20	8	2.0	63	4	US-10-424-599-225902
21	8	2.0	64	4	US-10-425-115-264714
22	8	2.0	66	4	US-10-424-599-177043
23	8	2.0	72	4	US-10-424-599-173970
24	8	2.0	77	4	US-10-425-115-240939
25	8	2.0	82	4	US-10-425-115-230667
26	8	2.0	84	4	US-10-425-115-226277
27	8	2.0	89	4	US-10-425-115-281398

ALIGNMENTS

RESULT 1

US-10-408-765A-1465
; Sequence 1465, Application US/10408765A
; Publication No. US20040101874A1

; GENERAL INFORMATION:

; APPLICANT: Ghosh, Soumitra S.

; APPLICANT: Fahy, Eoin D.

; APPLICANT: Zhang, Bing

; APPLICANT: Gibson, Bradford W.

; APPLICANT: Taylor, Steven W.

; APPLICANT: Glenn, Gary W.

; APPLICANT: Warnock, Dale E.

; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION

; IDENTIFIED IN THE MITOCHONDRIAL PROTEOME

; FILE REFERENCE: 660088.465

; CURRENT APPLICATION NUMBER: US/10/408,765A

; CURRENT FILING DATE: 2003-04-04

; NUMBER OF SEQ ID NOS: 3077

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1465

; LENGTH: 393

; TYPE: PRT

; ORGANISM: Homo sapiens

; US-10-408-765A-1465

Query Match 100.0%; Score 393; DB 4; Length 393;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGKSDFLTPKAIANRIKSGKGLKRLWYCMQCKQCRDENGFKCHCMSESHORQLLASEN 60

Db 1 MGKSDFLTPKAIANRIKSGKGLKRLWYCMQCKQCRDENGFKCHCMSESHORQLLASEN 60

Qy 61 PQQPMDFSEFRNDFLELLRRRFGTKRVHNNIYVNEYISHREHNNATOWETLTDTTK 120

Db 61 PQQPMDFSEFRNDFLELLRRRFGTKRVHNNIYVNEYISHREHNNATOWETLTDTTK 120

Qy 121 WLGRGLCKVDYTPKGYWYQVDRDPETIRRLQLEKKKQDLDEEKTAKFIEQVRRG 180

Db 121 WLGRGLCKVDYTPKGYWYQVDRDPETIRRLQLEKKKQDLDEEKTAKFIEQVRRG 180

Qy 181 LEGKEQVPTTELSENDEBEKVTFNLSKGACSSSGATSSKSTLGPALKTIIGSSASVK 240

Db 181 LEGKEQVPTTELSENDEBEKVTFNLSKGACSSSGATSSKSTLGPALKTIIGSSASVK 240

Qy 241 RKSSQSSTQSKKKKKKSSALDEIMEIEEKKRTARDYWLQPEIIVKIIITKKLGEKYHK 300

Db 241 RKSSQSSTQSKKKKKKSSALDEIMEIEEKKRTARDYWLQPEIIVKIIITKKLGEKYHK 300

Qy 301 KKAIVKEVIDKYTAVVQWIDSGDKLKDQTHLETVPAPGKRILVLNGVGRNGEGLTLESI 360

Db 301 KKAIVKEVIDKYTAVVQWIDSGDKLKDQTHLETVPAPGKRILVLNGVGRNGEGLTLESI 360

Db 301 KKAIVKEVIDKTVAVVYKMDISGDKLQDTHLETVPAPGKRILVNLGGYRGNEGTFLESI 360
QY 361 NEKTSATIVETGPKLGRVVEGIOVEDISKLA 393
Db 361 NEKTSATIVETGPKLGRVVEGIOVEDISKLA 393
RESULT 2
US-10-106-698-5464
; Sequence 5464, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5464
; LENGTH: 84
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-5464
Query Match 18.6%; Score 73; DB 4; Length 84;
Best Local Similarity 100.0%; Pred. No. 1.3e-59;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 321 SGDKLQDTHLETVPAPGKRILVNLGGYRGNEGTFLESI NEKTSATIVETGPKLGR 380
Db 12 SGDKLQDTHLETVPAPGKRILVNLGGYRGNEGTFLESI NEKTSATIVETGPKLGR 71
QY 381 VEGIOVEDISKLA 393
Db 72 VEGIOVEDISKLA 84
RESULT 3
US-10-767-701-57151
; Sequence 57151, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 57151
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 30963818.p
US-10-767-701-57151
Query Match 8.9%; Score 35; DB 4; Length 120;
Best Local Similarity 100.0%; Pred. No. 4.2e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHORQ 53

Db 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHORQ 53
RESULT 4
US-10-437-963-191916
; Sequence 191916, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 191916
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_88190C.1.p
US-10-437-963-191916
Query Match 8.9%; Score 35; DB 4; Length 138;
Best Local Similarity 100.0%; Pred. No. 4.7e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHORQ 53
Db 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHORQ 53
RESULT 5
US-10-739-930-8499
; Sequence 8499, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 8499
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: ZEAMA-23APR03-C7583_2.p
US-10-739-930-8499
Query Match 8.9%; Score 35; DB 5; Length 423;
Best Local Similarity 100.0%; Pred. No. 1.2e-23;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHORQ 53
Db 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHORQ 53
RESULT 6
US-10-425-115-341788
; Sequence 341788, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:

```
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 115
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 341788
; LENGTH: 424
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_74879C.1.pep
US-10-425-115-341788

Query Match      8.9%; Score 35; DB 4; Length 424;
Best Local Similarity 100.0%; Pred. No. 1.2e-23;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRYCQMCQKQCRDENGFKCHCMSESHQRQ 53
    |||||
DB 19 KGLQKLRYCQMCQKQCRDENGFKCHCMSESHQRQ 53
    |||||

RESULT 7
US-10-425-114-48183
; Sequence 48183, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 48183
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3591-011-E3_FLI.pep
US-10-425-114-48183

Query Match      8.9%; Score 35; DB 4; Length 437;
Best Local Similarity 100.0%; Pred. No. 1.3e-23;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRYCQMCQKQCRDENGFKCHCMSESHQRQ 53
    |||||
DB 32 KGLQKLRYCQMCQKQCRDENGFKCHCMSESHQRQ 66
    |||||

RESULT 8
US-10-425-114-58982
; Sequence 58982, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
```

```
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 58982
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700238628_FLI.pep
US-10-425-114-58982

Query Match      8.9%; Score 35; DB 4; Length 437;
Best Local Similarity 100.0%; Pred. No. 1.3e-23;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRYCQMCQKQCRDENGFKCHCMSESHQRQ 53
    |||||
DB 32 KGLQKLRYCQMCQKQCRDENGFKCHCMSESHQRQ 66
    |||||

RESULT 9
US-10-424-599-168326
; Sequence 168326, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 168326
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_123013C.1.pep
US-10-424-599-168326

Query Match      7.6%; Score 30; DB 4; Length 176;
Best Local Similarity 100.0%; Pred. No. 2.6e-19;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRYCQMCQKQCRDENGFKCHCMSE 48
    |||||
DB 19 KGLQKLRYCQMCQKQCRDENGFKCHCMSE 48
    |||||

RESULT 10
US-11-097-143-41040
; Sequence 41040, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
```

```
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41040
; LENGTH: 390
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-41040

Query Match          3.8%; Score 15; DB 6; Length 390;
Best Local Similarity 100.0%; Pred. No. 4.8e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 KSKGLQKLRWYCOMC 31
DB 17 KSKGLQKLRWYCOMC 31

RESULT 11
US-10-424-599-160516
; Sequence 160516, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 160516
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_115964C.1.pep
US-10-424-599-160516

Query Match          2.3%; Score 9; DB 4; Length 146;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 251 SKEKKKKKS 259
DB 56 SKEKKKKKS 64

RESULT 12
US-10-424-599-264493
; Sequence 264493, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
```

```
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 264493
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(374)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_80858C.1.pep
US-10-424-599-264493

Query Match          2.3%; Score 9; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 251 SKEKKKKKS 259
DB 124 SKEKKKKKS 132

RESULT 13
US-10-022-066-38
; Sequence 38, Application US/10022066
; Publication No. US20030166057A1
; GENERAL INFORMATION:
; APPLICANT: HILDEBRAND, WILLIAM H.
; APPLICANT: PRILLMAN, KILEY RAE
; TITLE OF INVENTION: METHOD AND APPARATUS FOR THE PRODUCTION OF ANTIGENS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 6680.034
; CURRENT APPLICATION NUMBER: US/10/022,066
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/256,410
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 60/256,409
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 09/465,321
; PRIOR FILING DATE: 1999-12-17
; PRIOR APPLICATION NUMBER: 09/974,366
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 638
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 38
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-022-066-38

Query Match          2.0%; Score 8; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.7e+06;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 SHORQLLL 56
DB 1 SHORQLLL 8

RESULT 14
US-10-022-066-610
; Sequence 610, Application US/10022066
; Publication No. US20030166057A1
; GENERAL INFORMATION:
; APPLICANT: HILDEBRAND, WILLIAM H.
; APPLICANT: PRILLMAN, KILEY RAE
; TITLE OF INVENTION: METHOD AND APPARATUS FOR THE PRODUCTION OF ANTIGENS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 6680.034
; CURRENT APPLICATION NUMBER: US/10/022,066
```


Wed Nov 30 09:40:51 2005

```
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/256,410
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 60/256,409
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 09/465,321
; PRIOR FILING DATE: 1999-12-17
; PRIOR APPLICATION NUMBER: 09/974,366
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 638
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 610
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: peptide
US-10-022-066-610

Query Match          2.0%; Score 8; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.7e+06;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      49 SHORQLLL 56
Db      1 SHORQLLL 8

RESULT 15
US-10-425-115-357220
; Sequence 357220, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 357220
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_88955C.1.pap
US-10-425-115-357220

Query Match          2.0%; Score 8; DB 4; Length 39;
Best Local Similarity 100.0%; Pred. No. 21;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      252 KERKKKKKS 259
Db      31 KERKKKKKS 38

Search completed: November 23, 2005, 16:55:49
Job time : 112.283 secs
```

This Page Blank (usp1c)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

QM protein - protein search, using sw model

Run on: November 23, 2005, 16:40:00 ; Search time 5.01276 Seconds
(without alignments)
237.628 Million cell updates/sec

Title: US-09-555-529-26

Perfect score: 393

Sequence: 1 MGKSDRLTPKAIANRIKSG.....GPLKGRVVEGIQYEDISKLA 393

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 17545 seqs, 3030971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 17545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA New:
1: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
4: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
5: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep.*
7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB pep.*
8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	2.0	120	1	US-10-821-234-1011
2	6	1.5	27	1	US-10-952-535A-31
3	6	1.5	57	1	US-10-723-207-88
4	6	1.5	112	1	US-10-793-626-2114
5	6	1.5	133	1	US-10-667-295-49
6	6	1.5	149	1	US-10-986-501-231
7	6	1.5	175	1	US-10-667-295-48
8	6	1.5	176	1	US-10-507-275-10
9	6	1.5	204	1	US-10-793-626-2394
10	6	1.5	226	1	US-10-485-517-186
11	6	1.5	227	1	US-10-986-501-136
12	6	1.5	229	1	US-10-793-626-2642
13	6	1.5	229	7	US-11-014-629-1
14	6	1.5	258	7	US-11-038-284-23
15	6	1.5	286	1	US-10-667-295-60
16	6	1.5	286	7	US-11-137-465-48
17	6	1.5	327	1	US-10-667-295-59
18	6	1.5	339	1	US-10-821-234-1507
19	6	1.5	389	7	US-11-012-762-72
20	6	1.5	398	7	US-11-012-762-74
21	6	1.5	441	1	US-10-510-386-162
22	6	1.5	458	7	US-11-077-550-114
23	6	1.5	467	1	US-10-510-386-56
24	6	1.5	471	1	US-10-467-962B-10
25	6	1.5	548	7	US-11-077-550-24

26 1.5 572 1 US-10-821-234-1290 Sequence 1290, Ap
27 6 600 1 US-10-131-826A-462 Sequence 462, App
28 6 674 1 US-10-507-275-9 Sequence 9, Appl
29 6 724 1 US-10-793-626-968 Sequence 968, App
30 6 749 1 US-10-793-626-138 Sequence 138, App
31 6 853 1 US-10-821-234-1110 Sequence 1110, Ap
32 6 858 7 US-11-077-550-22 Sequence 22, Appl
33 6 860 7 US-11-077-550-175 Sequence 175, App
34 6 862 7 US-11-077-550-94 Sequence 94, Appl
35 6 862 7 US-11-077-550-171 Sequence 171, App
36 6 862 7 US-11-077-550-173 Sequence 173, App
37 6 864 7 US-11-077-550-102 Sequence 102, App
38 6 865 7 US-11-077-550-100 Sequence 100, App
39 6 866 7 US-11-077-550-88 Sequence 88, Appl
40 6 866 7 US-11-077-550-104 Sequence 104, App
41 6 867 7 US-11-077-550-80 Sequence 80, Appl
42 6 867 7 US-11-077-550-96 Sequence 96, Appl
43 6 867 7 US-11-077-550-98 Sequence 98, Appl
44 6 870 7 US-11-077-550-92 Sequence 92, Appl
45 6 871 7 US-11-077-550-84 Sequence 84, Appl

ALIGNMENTS

RESULT 1

US-10-821-234-1011
; Sequence 1011, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821.234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 1011
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1011

Query Match 2.0%; Score 8; DB 1; Length 120;
Best Local Similarity 100.0%; Pred. No. 0.34;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 252 KEKKKKKS 259
Db 15 KEKKKKKS 22

RESULT 2

US-10-952-535A-31
; Sequence 31, Application US/10952535A
; Publication No. US20050255113A1
; GENERAL INFORMATION:
; APPLICANT: Huston, James S.
; APPLICANT: Messer, Anne
; APPLICANT: Lecerif, Jean-Michel
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITING POLYPEPTIDE-
; FILE REFERENCE: INR-004CP
; TITLE OF INVENTION: ACCUMULATION ASSOCIATED WITH NEUROLOGICAL DISORDERS
; CURRENT APPLICATION NUMBER: US/10/952.535A
; CURRENT FILING DATE: 2004-09-27
; PRIOR APPLICATION NUMBER: 60/146,047
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 45

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
; LENGTH: 27
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: construct
US-10-952-535A-31

Query Match          1.5%; Score 6; DB 1; Length 27;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 254 KKKKK 259
Db 15 KKKKK 20

RESULT 3
US-10-723-207-88
; Sequence 88, Application US/10723207
; Publication No. US20050250934A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; APPLICANT: Walfield, Alan M.
; TITLE OF INVENTION: PEPTIDE COMPOSITION AS IMMUNOGEN FOR THE TREATMENT OF
; TITLE OF INVENTION: ALLERGY
; FILE REFERENCE: 1151-4153US2
; CURRENT APPLICATION NUMBER: US/10/723,207
; CURRENT FILING DATE: 2003-11-24
; PRIOR APPLICATION NUMBER: 09/701,623
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US99/13959
; PRIOR FILING DATE: 1999-06-21
; PRIOR APPLICATION NUMBER: 09/100,287
; PRIOR FILING DATE: 1998-06-20
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 88
; LENGTH: 57
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Peptide
; OTHER INFORMATION: synthesized from amino acids with no genetic
; OTHER INFORMATION: material as source
US-10-723-207-88

Query Match          1.5%; Score 6; DB 1; Length 57;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 KKKKK 257
Db 11 KKKKK 16

RESULT 4
US-10-793-626-2114
; Sequence 2114, Application US/10793626
; Publication No. US2005025478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2114

; LENGTH: 112
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (112)
; OTHER INFORMATION: variable amino acid
US-10-793-626-2114

Query Match          1.5%; Score 6; DB 1; Length 112;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 QQLLL 56
Db 35 QQLLL 40

RESULT 5
US-10-667-295-49
; Sequence 49, Application US/10667295
; Publication No. US20050257293A1
; GENERAL INFORMATION:
; APPLICANT: Mascia, Peter
; TITLE OF INVENTION: BIOLOGICAL CONTAINMENT SYSTEM
; FILE REFERENCE: 11696-047001
; CURRENT APPLICATION NUMBER: US/10/667,295
; CURRENT FILING DATE: 2003-09-17
; PRIOR APPLICATION NUMBER: US 60/411,823
; PRIOR FILING DATE: 2002-09-17
; NUMBER OF SEQ ID NOS: 263
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49
; LENGTH: 133
; TYPE: PRT
; ORGANISM: Brassica napus
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(133)
; OTHER INFORMATION: Ceres Seq. ID no. 6416839
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(133)
; OTHER INFORMATION: Xaa = any amino acid
US-10-667-295-49

Query Match          1.5%; Score 6; DB 1; Length 133;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 SSASVK 240
Db 18 SSASVK 23

RESULT 6
US-10-986-501-231
; Sequence 231, Application US/10986501
; Publication No. US20050244845A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 90 Human Secreted Proteins
; FILE REFERENCE: P2013P2C1
; CURRENT APPLICATION NUMBER: US/10/986,501
; CURRENT FILING DATE: 2004-11-12
; PRIOR APPLICATION NUMBER: US/10/621,363
; PRIOR FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: 09/969,730
; PRIOR FILING DATE: 2001-10-06
; PRIOR APPLICATION NUMBER: 09/774,639
```

```
; PRIOR FILING DATE: 2001-02-01
; PRIOR APPLICATION NUMBER: 60/238,291
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 09/244,112
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: PCT/US98/16235
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 60/056,371
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,732
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,366
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,364
; PRIOR FILING DATE: 1997-08-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 373
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 231
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-986-501-231

Query Match      1.5%; Score 6; DB 1; Length 149;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 LEKKKK 160
DB 141 LEKKKK 146
|||||

RESULT 7
US-10-667-295-48
; Sequence 48, Application US/10667295
; Publication No. US20050257293A1
; GENERAL INFORMATION:
; APPLICANT: Mascia, Peter
; TITLE OF INVENTION: BIOLOGICAL CONTAINMENT SYSTEM
; FILE REFERENCE: 11696-047001
; CURRENT APPLICATION NUMBER: US/10/667,295
; CURRENT FILING DATE: 2003-09-17
; PRIOR APPLICATION NUMBER: US 60/411,823
; PRIOR FILING DATE: 2002-09-17
; NUMBER OF SEQ ID NOS: 263
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Brassica napus
; NAME/KEY: VARIANT
; LOCATION: (1)...(175)
; OTHER INFORMATION: Ceres Seq. ID no. 6416838
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(175)
; OTHER INFORMATION: Xaa = any amino acid
US-10-667-295-48

Query Match      1.5%; Score 6; DB 1; Length 175;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 SSASVK 240
DB 60 SSASVK 65
|||||

RESULT 8
US-10-507-275-10
; Sequence 10, Application US/10507275
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; Publication No. US20050250166A1
; GENERAL INFORMATION:
; APPLICANT: Masai, Hisao
; APPLICANT: Tamai, Katsuyuki
; APPLICANT: Medical and Biological Laboratories Co., Ltd.
; APPLICANT: Japan Science and Technology Agency
; APPLICANT: Ginkgo Biomedical Research Institute Co., Ltd.
; TITLE OF INVENTION: Cdc7-ASK Kinase Complex, Substrates of the Kinase Complex,
; TITLE OF INVENTION: Specific Antibodies to the Substrates, and Screening Methods
; TITLE OF INVENTION: Using the Same to Screen for Compounds Comprising Cdc7-ASK
; TITLE OF INVENTION: Kinase Inhibitory Ability
; FILE REFERENCE: 082368-0011000S
; CURRENT APPLICATION NUMBER: US/10/507,275
; CURRENT FILING DATE: 2004-09-09
; PRIOR APPLICATION NUMBER: JP 2002-067702
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: WO PCT/JP03/02918
; PRIOR FILING DATE: 2003-03-12
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-507-275-10

Query Match      1.5%; Score 6; DB 1; Length 176;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 KEKKKK 257
DB 115 KEKKKK 120
|||||

RESULT 9
US-10-793-626-2394
; Sequence 2394, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2394
; LENGTH: 204
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-10-793-626-2394

Query Match      1.5%; Score 6; DB 1; Length 204;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 QRQLLL 56
DB 35 QRQLLL 40
|||||

RESULT 10
US-10-485-517-186
; Sequence 186, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
```

```
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WO
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 186
; TYPE: PRT
; LENGTH: 226
; ORGANISM: Staphylococcus aureus
US-10-485-517-186

Query Match          1.5%; Score 6; DB 1; Length 226;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 232 TIGSSA 237
DB 124 TIGSSA 129

RESULT 11
US-10-986-501-136
; Sequence 136, Application US/10986501
; Publication No. US20050244845A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 90 Human Secreted Proteins
; FILE REFERENCE: P2013P2C1
; CURRENT APPLICATION NUMBER: US/10/986,501
; CURRENT FILING DATE: 2004-11-12
; PRIOR APPLICATION NUMBER: US/10/621,363
; PRIOR FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: 09/969,730
; PRIOR FILING DATE: 2001-10-06
; PRIOR APPLICATION NUMBER: 09/774,639
; PRIOR FILING DATE: 2001-02-01
; PRIOR APPLICATION NUMBER: 60/238,291
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 09/244,112
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: PCT/US98/16235
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 60/056,371
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,732
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,366
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,364
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 373
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 136
; TYPE: PRT
; LENGTH: 227
; ORGANISM: Homo sapiens
US-10-986-501-136

Query Match          1.5%; Score 6; DB 1; Length 227;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 253 EKKKKK 258
DB 124 TIGSSA 129

; APPLICANT: Kimmerly, William John
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P03480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2642
; LENGTH: 227
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-2642

Query Match          1.5%; Score 6; DB 1; Length 227;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 355 GTLESI 360
DB 57 GTLESI 62

RESULT 13
US-11-014-629-1
; Sequence 1, Application US/11014629
; Publication No. US20050244376A1
; GENERAL INFORMATION:
; APPLICANT: Hall, Frederick L.
; APPLICANT: Gordon, Erlinda M.
; TITLE OF INVENTION: TARGETED GENE DELIVERY TO EXTRACELLULAR MATRIX
; FILE REFERENCE: 30863-704.302
; CURRENT APPLICATION NUMBER: US/11/014,629
; CURRENT FILING DATE: 2004-12-15
; PRIOR APPLICATION NUMBER: US 08/837,223
; PRIOR FILING DATE: 1997-04-10
; PRIOR APPLICATION NUMBER: US 09/904,923
; PRIOR FILING DATE: 2001-07-13
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 229
; TYPE: PRT
; ORGANISM: moloney murine leukemia virus
US-11-014-629-1

Query Match          1.5%; Score 6; DB 7; Length 229;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 LKLDQT 330
DB 103 LKLDQT 108

RESULT 14
US-11-038-284-23
; Sequence 23, Application US/11038284
; Publication No. US20050246793A1
; GENERAL INFORMATION:
; APPLICANT: COOKE, DAVID
```

; APPLICANT: DEBET, MARTINE
 ; APPLICANT: GIDLEY, MICHAEL, JOHN
 ; APPLICANT: JOBLING, STEPHEN, ALAN
 ; APPLICANT: SAFFORD, RICHARD
 ; APPLICANT: SIDEBOTTOM, CHRISTOPHER, MICHAEL
 ; APPLICANT: WESTCOTT, ROGER, JOHN
 ; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO PLANT STARCH COMPOSITION
 ; FILE REFERENCE: 054163-5003-US
 ; CURRENT APPLICATION NUMBER: US/11/038,284
 ; CURRENT FILING DATE: 2005-01-21
 ; PRIOR APPLICATION NUMBER: US/10/056,454
 ; PRIOR FILING DATE: 2002-01-24
 ; PRIOR APPLICATION NUMBER: PCT/GB96/01075
 ; PRIOR FILING DATE: 1996-05-03
 ; PRIOR APPLICATION NUMBER: GB 9607409.1
 ; PRIOR FILING DATE: 1996-04-10
 ; PRIOR APPLICATION NUMBER: GB 9509229.2
 ; PRIOR FILING DATE: 1995-05-05
 ; NUMBER OF SEQ ID NOS: 43
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 23
 ; LENGTH: 258
 ; TYPE: PRT
 ; ORGANISM: Zea mays
 US-11-038-284-23

Query Match 1.5%; Score 6; DB 7; Length 258;
 Best Local Similarity 100.0%; Pred. No. 59;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 198 NDEEKV 203
 Db 101 NDEEKV 106

RESULT 15
 US-10-667-295-60
 ; Sequence 60, Application US/10667295
 ; Publication No. US20050257293A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Mascia, Peter
 ; TITLE OF INVENTION: BIOLOGICAL CONTAINMENT SYSTEM
 ; FILE REFERENCE: 11696,047001
 ; CURRENT APPLICATION NUMBER: US/10/667,295
 ; CURRENT FILING DATE: 2003-09-17
 ; PRIOR APPLICATION NUMBER: US 60/411,823
 ; PRIOR FILING DATE: 2002-09-17
 ; NUMBER OF SEQ ID NOS: 263
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 60
 ; LENGTH: 286
 ; TYPE: PRT
 ; ORGANISM: Arabidopsis thaliana
 ; FEATURE:
 ; NAME/KEY: VARIANT
 ; LOCATION: (1)...(286)
 ; OTHER INFORMATION: Ceres Seq. ID no. 13491862
 US-10-667-295-60

Query Match 1.5%; Score 6; DB 1; Length 286;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 235 SSASVK 240
 Db 21 SSASVK 26

Search completed: November 23, 2005, 16:52:01
 Job time : 5.01276 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:39:45 ; Search time 25.0638 Seconds
(without alignments)
1508.679 Million cell updates/sec

Title: US-09-555-529-26
Perfect score: 393
Sequence: 1 MGKSDFLTPKAIANRIKSG.....GPLKGRRVGIQYEDISKLA 393

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- PIR 80:*
- 1: Pir1.*
- 2: Pir2.*
- 3: Pir3.*
- 4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	163	41.5	391	2	S18666
2	35	8.9	411	2	H96596
3	10	2.5	404	2	T27106
4	9	2.3	304	2	T40316
5	8	2.0	178	2	G75018
6	8	2.0	178	2	A71205
7	8	2.0	303	1	S75983
8	8	2.0	521	2	B64181
9	8	2.0	646	2	F71620
10	8	2.0	1799	2	A1895
11	7	1.8	57	2	A12525
12	7	1.8	68	2	E31176
13	7	1.8	70	2	E81029
14	7	1.8	89	2	AF2709
15	7	1.8	89	2	G97491
16	7	1.8	94	2	F86022
17	7	1.8	100	2	D82245
18	7	1.8	106	2	S9536
19	7	1.8	114	2	S64961
20	7	1.8	115	2	T33703
21	7	1.8	121	2	D75089
22	7	1.8	132	2	T49635
23	7	1.8	133	2	T22262
24	7	1.8	136	2	AB2542
25	7	1.8	137	2	S37848
26	7	1.8	138	2	B84793
27	7	1.8	138	2	T45905
28	7	1.8	144	2	G65255
29	7	1.8	150	2	H82341

30	7	1.8	153	2	S59587
31	7	1.8	153	2	S59125
32	7	1.8	153	2	S59591
33	7	1.8	175	2	T05669
34	7	1.8	177	2	PS0374
35	7	1.8	206	2	T25384
36	7	1.8	214	2	JC7297
37	7	1.8	215	2	AB3358
38	7	1.8	217	2	JC7997
39	7	1.8	237	2	T27198
40	7	1.8	239	2	C84994
41	7	1.8	239	2	AF0780
42	7	1.8	250	2	T44896
43	7	1.8	250	2	D70787
44	7	1.8	253	2	A05283
45	7	1.8	266	2	A86288

ALIGNMENTS

RESULT 1

S18666
KIN17 protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C:Accession: S18666; A43753; S14622
R:Angulo, J.F.; Rouer, E.; Marin, A.; Mattei, M.G.; Tissier, A.; Horellou, P.; Benarous, N.
Nucleic Acids Res. 19, 5117-5123, 1991
A:Title: Identification and expression of the cDNA of KIN17, a zinc-finger gene located .
A:Reference number: S18666; MUID:92020193; PMID:1923796
A:Accession: S18666
A:Molecule type: mRNA
A:Residues: 1-391 <ANG>
A:Cross-references: UNIPROT:Q8K339; UNIPARC:UPI0000017993F; EMBL:X58472
R:Angulo, J.; Rouer, E.; Benarous, R.; Devoret, R.
Biochimie 73, 251-256, 1991
A:Title: Identification of a mouse cDNA fragment whose expressed polypeptide reacts with
A:Reference number: A43753; MUID:91355299; PMID:1715759
A:Accession: A43753
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 74-273 <ANW>
A:Cross-references: UNIPARC:UPI00000179940
C:Superfamily: KIN17 protein

Query Match	41.5%	Score 163;	DB 2;	Length 391;
Best Local Similarity	100.0%;	Pred. No. 1.5e-153;		
Matches 163;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	25	RWYCOMCOKCRDENGFKCHMSHQRQLLASENQQQPMDYFSEFRNDFLLELRFF	84	
Db	25	RWYCOMCOKCRDENGFKCHMSHQRQLLASENQQQPMDYFSEFRNDFLLELRFF	84	
Qy	85	GTKRVHNNIVYVNIYSHREIHNNATQWETLDTFKWLGREGLCCKVDTPKGVYQYIDR	144	
Db	85	GTKRVHNNIVYVNIYSHREIHNNATQWETLDTFKWLGREGLCCKVDTPKGVYQYIDR	144	
Qy	145	DPETIRQLELEKKKKQQLDDEKTAKEFIEQVRRGLEKQEQ	187	
Db	145	DPETIRQLELEKKKKQQLDDEKTAKEFIEQVRRGLEKQEQ	187	

RESULT 2

H96596
hypothetical protein T5A14.13 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: H96596
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federopiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: H96596
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-411 <STO>
A;Cross-references: UNIPROT:Q9ZVU5; UNIPARC:UPI00000A3CC8; GB:AE005173; NID:g4204268; P
C;Genetics:
A;Gene: TSA14.13
A;Map position: 1
C;Superfamily: KIN17 protein

Query Match 8.9%; Score 35; DB 2; Length 411;
Best Local Similarity 100.0%; Pred. No. 2.2e-26;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGIQKLRLWYQCMQKQCRDENGFKCHCMSESHQRQ 53
|||||
Db 19 KGIQKLRLWYQCMQKQCRDENGFKCHCMSESHQRQ 53

RESULT 3
T27106
hypothetical protein Y52B11A.9 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
R;Lennard, N.
Submitted to the EMBL Data Library, October 1998
A;Reference number: Z20310
A;Accession: T27106
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-404 <WIL>
A;Cross-references: UNIPROT:Q9XWF2; UNIPARC:UPI0000078B04; EMBL:AL032654; PIDN:CAA21720.
A;Experimental source: clone Y52B11A
C;Genetics:
A;Gene: CESP:Y52B11A.9
A;Map position: 1
A;Intons: 26/2; 102/3; 301/3; 367/3
C;Superfamily: KIN17 protein

Query Match 2.5%; Score 10; DB 2; Length 404;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 CQMCQKQCRD 37
|||||
Db 28 CQMCQKQCRD 37

RESULT 4
T40316
conserved kin7-like hypothetical protein SPBC365.09c - fission yeast (*Schizosaccharomyce*
C;Species: *Schizosaccharomyces pombe*
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T40316
R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Lelaure, V.; Galibert, F.
Submitted to the EMBL Data Library, June 1999
A;Reference number: Z21920
A;Accession: T40316
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-304 <WOO>
A;Cross-references: UNIPROT:Q9Y7X9; UNIPARC:UPI000006B457; EMBL:AL078627; PIDN:CAB44761.
A;Experimental source: strain 972h-; cosmid c365
C;Genetics:
A;Gene: SPDB:SPBC365.09c

A;Map position: 2

Query Match 2.3%; Score 9; DB 2; Length 304;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 RDENGFKCH 44
|||||
Db 36 RDENGFKCH 44

RESULT 5
F75018
inorganic diphosphatase (EC 3.6.1.1) PAB1104 [similarity] - *Pyrococcus abyssi* (strain Orf
C;Species: *Pyrococcus abyssi*
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
C;Accession: F75018
R;Anonymous, Genoscope
Submitted to the EMBL Data Library, July 1999
A;Description: *Pyrococcus abyssi* genome sequence: insights into archaeal chromosome stru
A;Reference number: A75001
A;Accession: F75018
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-178 <KAW>
A;Cross-references: UNIPROT:Q9UY24; UNIPARC:UPI000003456C; GB:AJ248288; GB:AL096836; NID:
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: ppa; PAB1104
C;Superfamily: inorganic pyrophosphatase
C;Keywords: hydrolase

Query Match 2.0%; Score 8; DB 2; Length 178;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 317 KNIDSGDK 324
|||||
Db 95 KNIDSGDK 102

RESULT 6
A71205
inorganic diphosphatase (EC 3.6.1.1) PH1907 [similarity] - *Pyrococcus horikoshii*
C;Species: *Pyrococcus horikoshii*
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 05-Oct-2004
C;Accession: A71205
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekiri
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: A71205
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-178 <KAW>
A;Cross-references: UNIPROT:O59570; UNIPARC:UPI0000052CB5; GB:AP000007; NID:g3236134; Pfi
A;Experimental source: strain Or3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH1907
C;Superfamily: inorganic pyrophosphatase
C;Keywords: hydrolase

Query Match 2.0%; Score 8; DB 2; Length 178;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 317 KNIDSGDK 324
|||||
Db 95 KNIDSGDK 102

```

RESULT 7
S75983
Hypothetical protein - Synochocystis sp. (strain PCC 6803)
C:Species: Synochocystis sp.
A:Variety: PCC 6803
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: S75983
R:Kaneko, T.; Sato, S.; Kofani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synochocystis
s.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S75983
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-303 <KAN>
A:Cross-references: UNIPROT:Q55483; UNIPARC:UPI000000COEFA; EMBL:D64006; GB:AB001339; NID
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: hypothetical protein ybbp

Query Match 2.0%; Score 8; DB 1; Length 303;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 76 FLELLRRR 83
|||||
Db 100 FLELLRRR 107

RESULT 8
E64181
Probable cytochrome oxidase (EC 1.10.3.-) chain I - Haemophilus influenzae (strain Rd KW
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 09-Jul-2004
C:Accession: E64181
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: E64181
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-521 <TIGR>
A:Cross-references: UNIPROT:P45021; UNIPARC:UPI0000128C53; GB:U32787; GB:L42023; NID:GL5
C:Superfamily: cytochrome d complex terminal oxidase chain I
C:Keywords: electron transfer; heme; oxidoreductase; transmembrane protein
F:20-36/Domain: transmembrane #status predicted <TM1>
F:100-116/Domain: transmembrane #status predicted <TM2>
F:192-208/Domain: transmembrane #status predicted <TM3>
F:222-238/Domain: transmembrane #status predicted <TM4>
F:398-414/Domain: transmembrane #status predicted <TM5>
F:426-442/Domain: transmembrane #status predicted <TM6>
F:477-493/Domain: transmembrane #status predicted <TM7>

Query Match 2.0%; Score 8; DB 2; Length 521;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 225 LGPSALKT 232
|||||
Db 504 LGPSALKT 511

RESULT 9
F71620
Hypothetical protein PFB0230c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 09-Jul-2004
C:Accession: F71620

```

```

R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.)
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: F71620
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-646 <GAR>
A:Cross-references: UNIPROT:O96147; UNIPARC:UPI000017B603; GB:AE001381; GB:AE001362; NID:
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0230c

Query Match 2.0%; Score 8; DB 2; Length 646;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 251 SKEKKKKK 258
|||||
Db 575 SKEKKKKK 582

RESULT 10
AD1895
serine/threonine kinase with two-component sensor domain alr0710 [imported] - Nostoc sp.
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C:Accession: AD1895
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD1895
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1799 <KUR>
A:Cross-references: UNIPROT:O8VYV2; UNIPARC:UPI000000CDB39; GB:BA000019; PIDN:BA075667.1
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr0710

Query Match 2.0%; Score 8; DB 2; Length 1799;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 213 SSSGATSS 220
|||||
Db 1325 SSSGATSS 1332

RESULT 11
AI2525
transposase asr7385 [imported] - Nostoc sp. (strain PCC 7120) plasmid pcc7120alpha
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C:Accession: AI2525
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AI2525
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-57 <KUR>
A:Cross-references: UNIPROT:O8VKB6; UNIPARC:UPI000000CEFC; GB:BA000020; PIDN:BA077143.1
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: asr7385

```

```

A:Genome: plasmid

Query Match      1.8%; Score 7; DB 2; Length 57;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 LLRRRFG 85
    |||||
Db 40 LLRRRFG 46

RESULT 12
E91176
hypothetical protein ECs4381 [imported] - Escherichia coli (strain O157:H7, substrain R1
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: E91176
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
; Sasawara, N.; Yasunaga, T.; Kihara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: E91176
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-68 <HAY>
A:Cross-references: UNIPROT:Q8X2W1; UNIPARC:UPI00000D2A54; GB:BA000007; PIDN:BA037804.1;
A:Experimental source: strain O157:H7, substrain R1MD 050952
C:Genetics:
A:Gene: ECs4381

Query Match      1.8%; Score 7; DB 2; Length 68;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 250 OSKEKKK 256
    |||||
Db 2 OSKEKKK 8

RESULT 13
E81029
hypothetical protein NMB1892 [imported] - Neisseria meningitidis (strain MC58 serogroup
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C:Accession: E81029
R:Tetzelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.
; Ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: E81029
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-70 <TET>
A:Cross-references: UNIPROT:Q9JXT4; UNIPARC:UPI00000C4805; GB:AE002539; GB:AE002098; NID
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1892

Query Match      1.8%; Score 7; DB 2; Length 70;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 RRGLECK 184
    |||||
Db 53 RRGLECK 59

RESULT 14
AF2709
hypothetical protein Atul079 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AF2709
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
; erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AF2709
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-89 <KUR>
A:Cross-references: UNIPROT:Q8UGF8; UNIPARC:UPI00000D1A2B; GB:AE008688; PIDN:AAL42092.1;
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atul079
A:Map position: circular chromosome

Query Match      1.8%; Score 7; DB 2; Length 89;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 KFIEQV 177
    |||||
Db 12 KFIEQV 18

RESULT 15
G97491
hypothetical protein AGR_C_1994 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C:Accession: G97491
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: G97491
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-89 <KUR>
A:Cross-references: UNIPROT:Q8UGF8; UNIPARC:UPI00000D1A2B; GB:AE007869; PIDN:AAK86888.1;
A:Genetics:
A:Gene: AGR_C_1994
A:Map position: circular chromosome

Query Match      1.8%; Score 7; DB 2; Length 89;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 KFIEQV 177
    |||||
Db 12 KFIEQV 18

Search completed: November 23, 2005, 16:51:47
Job time : 26.0638 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

QM protein - protein search, using sw model

Run on: November 23, 2005, 16:19:54 ; Search time 25.0638 Seconds
(without alignments)
1508.679 Million cell updates/sec

Title: US-09-555-529-26

Perfect score: 2037

Sequence: 1 MGKSDFLTPKAIANRIKSG.....GPLKGRVGEIQVEDISKLA 393

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_80.*

1: Pirl.*

2: Pirl.*

3: Pirl.*

4: Pirl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1887	92.6	391	2	S18666
2	978	48.0	411	2	H96596
3	952.5	46.8	1404	2	T27106
4	586.5	28.8	304	2	T40316
5	300.5	14.8	232	2	S50239
6	165	8.1	462	2	A86459
7	143	7.0	482	2	T24518
8	141.5	6.9	517	1	A39038
9	140	6.9	375	2	T05523
10	139.5	6.8	764	2	A84671
11	138	6.8	805	2	E70474
12	133.5	6.6	630	2	H96770
13	133	6.5	1359	2	T34036
14	131.5	6.5	1877	2	T21861
15	129.5	6.4	2401	2	T28676
16	128	6.3	741	2	S12921
17	127.5	6.3	2541	2	T29340
18	127	6.2	2094	2	S33124
19	127	6.2	3488	2	T34418
20	126.5	6.2	485	2	S44819
21	126.5	6.2	1827	2	T16270
22	126	6.2	671	2	H64502
23	126	6.2	888	2	E71608
24	125.5	6.2	978	2	A70387
25	125	6.1	1188	2	E89896
26	123.5	6.1	1359	2	T52091
27	123.5	6.1	1388	2	T38720
28	123	6.0	377	2	S52537
29	123	6.0	1164	2	T24806

ALIGNMENTS

RESULT 1

S18666

KIN17 protein - mouse

C;Species: Mus musculus (house mouse)

C;Date: 19-Feb-1994 #sequence revision 10-Nov-1995 #text_change 09-Jul-2004

C;Accession: S18666; A43753; S14622

R;Angulo, J.F.; Rouer, E.; Mazin, A.; Mattei, M.G.; Tisser, A.; Horellou, P.; Benarous;

Nucleic Acids Res. 19, 5117-5123, 1991

A;Title: Identification and expression of the cDNA of KIN17, a zinc-finger gene located

A;Reference number: S18666; MUID:92020193; PMID:1923796

A;Accession: S18666

A;Molecule type: mRNA

A;Residues: 1-391 <ANG>

A;Cross-references: UNIPROT:Q8K339; UNIPARC:UPI000017993F; EMBL:X58472

R;Angulo, J.; Rouer, E.; Benarous, R.; Devoret, R.

Biochimie 73, 251-256, 1991

A;Title: Identification of a mouse cDNA fragment whose expressed polypeptide reacts with

A;Reference number: A43753; MUID:91355299; PMID:1715759

A;Accession: A43753

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 74-273 <ANW>

A;Cross-references: UNIPARC:UPI0000179940

C;Superfamily: KIN17 protein

Query Match 92.6%; Score 1887; DB 2; Length 391;

Best Local Similarity 91.9%; Pred. No. 7.3e-100;

Matches 361; Conservative 15; Mismatches 15; Indels 2; Gaps 1;

Qy 1 MGKSDFLTPKAIANRIKSGLQKLRWYCOMQOCQCRDENGPKCHCMSESHORQLLASEN 60

Db 1 MGKSDFLTPKAIANRIKSGLQKLRWYCOMQOCQCRDENGPKCHCMSESHORQLLASEN 60

Qy 61 PQQFMDYFSEFRNDFLELLRRRFGTKRVHNNIYVYNEIYSHREIHNNATQWETLTDFTK 120

Db 61 PQQFMDYFSEFRNDFLELLRRRFGTKRVHNNIYVYNEIYSHREIHNNATQWETLTDFTK 120

Qy 121 WLGRGLCKVDTPKGYIYQYIDRDPETIRQLLEKKKKODLDEBKTAKFIEQVRRG 180

Db 121 WLGRGLCKVDTPKGYIYQYIDRDPETIRQLLEKKKKODLDEBKTAKFIEQVRRG 180

Qy 181 LEGKEQVPFTELSRNDEEKVTFNLSKGACSSSGATSSKSSLTGFSALKTICSSASVK 240

Db 181 LEGKEQVPFTELSRNDEEKVTFNLSKGACSSSGATSSKSSLTGFSALKTICSSASVK 240

Qy 241 RKSSQSSSTOSKSKKKKKSALDEIMEIEEKKRTARTDYMLQPIIYVKIITKLGKGYHK 300

Db 241 RKSSQSSSAQ - PAKKKKSALDEIMEIEEKKRTARTDNLQPGIVVKIITKLGKGYHK 298

Qy 301 KKAIVKEVIDKYTAVVKXIDSGDKLQDQTHLETVTPAPGKRILVNLGVNGVNGSGTLES 360

Db 299 KKGVVKEVIDRYTAVVKWMTDSDGLKLDQDTHLETVTPAPGKRVLVNLGVNGVNGSGTLES 358

myosin-like protei
centromere protein
tropoin T - fruit
membrane protein p
chromosome segrega
hypothetical prote
probable chromosom
h-caldesmon - chic
caldesmon - human
hypothetical prote
hypothetical prote
19A-specific metal
probable erythrocy
myosin heavy chain
nuclear histone-bi
hypothetical prote

A; Experimental source: clone T05D4
C; Genetics:
A; Gene: CESP.T05D4.5
A; Map position: 3
A; Introns: 128/3; 184/3; 212/3; 236/3; 305/3

Query Match 7.0%; Score 143; DB 2; Length 482;
Best Local Similarity 21.4%; Pred. No. 0.53;
Matches 87; Conservative 50; Mismatches 134; Indels 136; Gaps 18;

QY 2 GKSDFLTPK--AIANRIKSGLQLRWYC-----OMCOKQCRDENGFKCH-----44
DB 118 GKCFPFVDQRILIVGRGVEPVQ--HWFAIYNTSPFPYMFKIKCSDNSLFKLIHPVCGVL 175
QY 45 -CMSEHQRLLASENPQQFMDYFS-----EFRNDFLELLRRRFGTKRV 89
DB 176 DALCDAEISVTFCGAHVPPRYTSLSHSMLPDPLFLDVPOAFRN-----RRNFETRQ- 228
QY 90 HNNIVNYEYISHREHIHNAQTOWETLTDTFKWLGREGLCKVKDETPKGWYIQYIDR-----144
DB 229 ---RLVVEYAYHEDPI-----LVTAEEDRDQFHFYNRKVLKTD 264
QY 145 DPTIRROELEKKKKDKOLD-DDEKTAKFTIEQVRRGLEGKEQEVPTFELS-----196
DB 265 EASEPAKKKAELKVLRDKIDEIKARKIDEKUIER-EIVRAESPKFEPKTRRLRAVEKF 322
QY 197 -----END-BEKVTFNLSK-----GACSSSGATSSKSTLG 226
DB 323 FNQOKQKENWEKLKVDASRPMLSHDWLMKNIAEKKKIGAVKKKIKAEEKDDKK 382
QY 227 P----SALKTTIGSSASVSRKSSQSOSQSKBKKKKKXSALD-EIMETEIEKKRTARTDYWL 281
DB 383 PMGWMTLKWERDERRTKRKEEBEKKKDAEEKKKDAEKKKDAEKKKDAEGD---439
QY 282 QPEIIVKIIITKLGEXYHKKAIVKEVIDKYTAVVMIDSGDKLKLD 328
DB 440 -----KKXDAEDEKKKDAGDE-----KKKKDAGEKKKKD 468

RESULT 8
A39038
N; Alternate names: caldesmon, nonmuscle
C; Species: Gallus gallus (chicken)
C; Date: 31-Jul-1991 #sequence_revision 27-Jun-1994 #text_change 22-Jun-1999
C; Accession: A39038
R; Hayashi, K.; Fujio, Y.; Kato, I.; Sobue, K.
J. Biol. Chem. 266, 355-361, 1991
A; Title: Structural and functional relationships between h- and l-caldesmons.
A; Reference number: A39038; MUID: 91093148; PMID: 1824698
A; Accession: A39038
A; Molecule type: mRNA
A; Residues: 1-517 <HAY>
A; Cross-references: UNIPARC:UPT0000156296; GB:M60620; GB:M38015; NID:G212242; PIDN:AAA48
A; Experimental source: brain
C; Comment: The binding of caldesmon to F-actin is modulated by calcium and calmodulin.
C; Comment: Two calmodulin molecules can bind to nonoverlapping domains of each caldesmon
C; Superfamily: caldesmon
C; Keywords: actin binding; calmodulin binding; phosphoprotein
F; 342,427,462/Binding site: phosphate (Ser) (covalent) #status predicted
F; 433,456/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 6.9%; Score 141.5; DB 1; Length 517;
Best Local Similarity 20.9%; Pred. No. 0.69;
Matches 66; Conservative 58; Mismatches 105; Indels 87; Gaps 12;

QY 48 ESHQRQLLASNPQQFMDYFSBEFRNDFLELLRRFGTKRVHNNIVNYEYISHREHHM 107
DB 90 ERQKRLQEALERQKEFDPIITDGSL-----VPSREVNVNEEITKEEV--138
QY 108 NATQWETLTDTFKWLGREGLCKVKDETPKGWYIQYIDRDPETIR-----ROLEFKKK 159
DB 139 -----ET-----RQRCETIEET-----ETVTKSYORNNWRQGEGBEGK 171

Qy	160	KODLDDSEKTA	KFI---BEQVR-----RGLGKGQEVPTFTTSLRSNDEEKV	203
		:: ::	:: ::	
Db	172	KEEKDSESEKPE	VPTTEENQYKDNKVKEKAPKEEMKSWDRKRGVP---EQKAQNGEREL	228
		:: ::	:: ::	
Qy	204	T---FNLSKGAC	SSSGATSSKSSITGLPSALKTIGSSASV-----KKKSSQSSTQSKEK	254
		:: ::	:: ::	
Db	229	TTPKLKSTENAF	GRSNLKGANAAGSEKLEKEQOEAAVELDELKKRREPRKKILSEEQ	298
		:: ::	:: ::	
Qy	255	KKKSKALDEIM	EIEEEKRKTARTDYWLQPEIIVKIITKGLGEKYHKKAIVKSEVIDKYTA	314
		:: ::	:: ::	
Db	289	KKKQSEAEKRI	EEBEKKR-----MKEEIERRAAEAAKKQKQVPEDGVSEKK	336
		:: ::	:: ::	
Qy	315	VVKMID-SGD	KLKLDQ	329
		:: ::	:: ::	
Db	337	PPKCFSPKXG	SSLKITE	352
		:: ::	:: ::	

RESULT 9

T05523

hypothetical protein F13M23.160 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004

C;Accession: T05523

R;Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hohseisel, J.; Mewes, H.W.; M

submitted to the Protein Sequence Database, February 1999

A;Reference number: Z15419

A;Accession: T05523

A;Molecule type: DNA

A;Residues: 1-375 <REV>

A;Cross-references: UNIPROT:Q9SW24; UNIPARC:UPI000000AGF0F; EMBL:AL035523

C;Genetic:

A;Map position: 4

A;Note: F13M23.160

Query Match	6.9%;	Score 140;	DB 2;	Length 375;
Best Local Similarity	23.5%;	Pred. No. 0.58;		
Matches	64;	Conservative	47;	Mismatches 99; Indels 62; Gaps 10;
Qy	132	ETPKGWYQIYIDRDPETIRRQLELEKKKKQDLDDDEKTAKEEQQVRRGLE----	GKEQE	187
Db	80	EIPSG-----DGSNNITVGLNL-RQKVENPEPEPEKMLLKSMTDLESPLDAPEKE	132	
Qy	188	-----VPTFTE-----LSRENDEEKVTNLSKGACSSSSGSSKSTLGG	226	
Db	133	DFESFPVDGFEALLAGYGMWPGGIGLKAKEDVKIYEYKWSGNEGFGFGKSDKAWMI	192	
Qy	227	PSALKTIGSSAYK-----RKSSQSSTOSKEKKKKKXALDEIMEI	267	
Db	193	DNKLVGSGSHEEVKVGINKIENMEKDRVVRKRNRETEGESRTVEKACKQNYRG-----QT	247	
Qy	268	EEBKKTARDYDYLQPEIIVKIITKL-GBKYHKKAIVKEVIDKYTAVVKMIDSGDKLK	326	
Db	248	RETREKTS-----WRSHKIVRIISKDVKGGRLYLKAAVTDVVVGPTSCDIAMDETQLVQ	303	
Qy	327	LDQTHLETVIPAGKRILVINGYRGNECTL	357	
Db	304	GIDOELELTALPRGGSVLVLSCRHKGVYGRLL	335	

RESULT 10

A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.
 A:Reference number: A84420; MUID:20083487; PMID:10617197
 Nature 402, 761-768, 1999
 C:Accession: A84671
 R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii,
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tal-
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
 C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
 C/Species: *Arabidopsis thaliana* (mouse-ear cress)
 C/Hypothetical protein At3g27280 [imported] - *Arabidopsis thaliana*
 A84671

A:Accession: A84671
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-764 <STO>
C:Cross-references: UNIPROT:Q9XINS; UNIPARC:UPI000017A6B9; GB:AE002093; NID:g5306269; P1
A:Gene: At2g27280
A:Map position: 2

Query Match 6.8%; Score 139.5; DB 2; Length 764;
Best Local Similarity 19.0%; Pred. No. 1.4;
Matches 78; Conservative 77; Mismatches 134; Indels 121; Gaps 17;

QY 46 MSSEHORQLLASNPQF-----MDYFSEFRNDFLELLRRRFGTK 87
DB 54 IESQHKAL--BEDPSAFSYDEVYDDMKQKAVLPRMQD--REERKPYIQLMKQARR 108
QY 88 RVHNNIVYNEIYHGR-----EHIHNNATQMET-----LTDFTKWLGRGLCKVDETPKGW 137
DB 109 EKEHEIVYERKLAKEREKDEHLFSDEKFTVGAYKRKLEBQKWLAEERLRE-----163
QY 138 YIOYIDRDPETIRRL-----ELEKKKQDLDEKTAKFIEBQVRRGL 181
DB 164 ----ERDDVTKKDLSDFFYNIGKNVAFGAREVEAKEAEKLEBQKAEKLEBQKAEKL 218
QY 182 EGKEQEVPTTFLSRENDEKVTFNLSKGACSSSGATSSKSTLGP-----SALKTIGS 235
DB 219 EELRKEY-TRVEKKRSPEKEV-----SPDGEFGSSRSKSLPLEABQAVSEKMG 270
QY 236 SASVKRESSQSSTQS-----KEKKKKSALDEME-----TEEEKKRT 274
DB 271 DGTERRSSKIEAKEVPKAINDKREDIAAAKERFLARSTISITWAILMKNKENEQ 330
QY 275 ARTDY-----WLOPEIIVKIITKLGK--YHKKKAIVKEVIDKYTA 314
DB 331 RKIPYVEMHIVRVSLNVLVDQIEWYIGKRFLENVLDPFGKESNFKLRTAKESITDAVTT 390
QY 315 VVKMIDSGDKLDQTHLETVIPAGKRIILVNLGGYRGNE--GTLESINE 362
DB 391 AKERLRETER-KTESMDVEKVRPS-----TLFPNASFDPSDPLGFLKVPF 435

RESULT 11
E70474
translation initiation factor IF-2 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
C:Accession: E70474
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: E70474
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-805 <AQF>
A:Cross-references: UNIPROT:O67825; UNIPARC:UPI0000056795; GB:AE000769; NID:g2984262; P1
A:Experimental source: strain VF5
C:Genetics:
A:Gene: infB
C:Superfamily: translation initiation factor IF-2; translation elongation factor Tu homC
F:Keywords: GTP binding; nucleotide binding; P-loop
F:305-419/Domain: translation elongation factor Tu homology <ETU>
F:315-322/Region: nucleotide-binding motif A (P-loop)
F:416-419/Region: GTP-binding NKXD motif
F:452-454/Region: GTP-binding SAK/L motif
F:321,322,342,416,417,419,452/Binding site: Mg-GTP (Lys, Thr, Thr, Asn, Lys, Asp, Ser) #

Query Match 6.8%; Score 138; DB 2; Length 805;
Best Local Similarity 22.4%; Pred. No. 1.8;
Matches 75; Conservative 54; Mismatches 106; Indels 100; Gaps 17;

QY 143 DRDPETIRQLEKKKKQDLDDEKTAKFIEBQVRRGLGKQEVPTTFLSRENDEK 202
DB 99 EKKPEVIVEIEB-EKKEBEKKEEKPKSVSEELIKLEKKEKEK-KKVEKERKEK 156
QY 203 VTFNLSKGACSSSGATSSKSTLGPALKTISSASVKR-----KSSQSSTQSKEKK 255
DB 157 VRVVEVKKEERKEEKEEKEPK-KIWSKKREIMRKLHAEVEKKEKKOKEKKEKK 215
QY 256 KKKS-----ALDEIMEIEEBK-----KR--TARTDYWLOPEIIVKIITK- 292
DB 216 KKEEVEKIIYDEVITVRELAELODPANKVIAELMKRGVLATINOPVPEVAEVSF 275
QY 293 -----KLGEYHKKKKAIKVIDK-----YTAIVKMIDSGDKLKDQTHLETV-- 335
DB 276 GYLAEVKKEEELKEEALLKEEEREELQPRPPIVVVMGHVDHGKTTLLDRIRKTNVAE 335
QY 336 -----TPAP-GKRILVLN-CGY-----RG-----NEG--- 355
DB 336 REKGGITOHIGASQVELPDGRKITPLDTPGHEAFTTLRARGAKVTDISVLVWAADGVMP 395
QY 356 -TLESINE-KTFSATIVI-----ETGPLKGR 380
DB 396 QTIEAINHAKAFNPVPIIVAVNKIDKPNADPMKVR 430

RESULT 12
H96770
protein heat shock protein P1017.8 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: H96770
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federapfel, N.A.; Kaul, S.; White, O.; Alonso;
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: H96770
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-630 <STO>
A:Cross-references: UNIPROT:Q9C911; UNIPARC:UPI000000AA4E1; GB:AE005173; NID:g6939229; P1
C:Genetics:
A:Gene: P1017.8
A:Map position: 1

Query Match 6.8%; Score 133.5; DB 2; Length 630;
Best Local Similarity 21.1%; Pred. No. 2.5;
Matches 74; Conservative 54; Mismatches 133; Indels 89; Gaps 13;

QY 27 YCQWCKQCRDENGFKCHCSHSHORLLASNPQFMDYFSEFRNDFLELLRRRFGT 86
DB 309 YCIVCSKKFKSEKQWKHQSCKHKEKV---AELRESFTDYEEENEEDIDGPLDSPEV 365
QY 87 KRVHNNIVYNEYISHREIHNNATQMETLDTFTKWLGRGLCKVDETPKGWYI----- 139
DB 366 BELHEKLOEELNIDNEB-----RDVKKEVVGEADETDDEYFAEEDMQG 409
QY 140 --QVIDRDPETIRQLEKKKKQ-----DLDEKTAKFIE-----EQVRR 179
DB 410 SSESDEDDMTLLKKVSGQKNQKNVSKKEDEDETEVEIGDTAFSEFDFNOKSTGR 469
QY 180 GLECKEVEPTTFLSRENDEKVTFNLSKGACSSSG-----ATSSKSTLGPALK 231
DB 470 NKEAKE-----ERNKQAGNMDADDTSKVQIPGCGNPDENNNATESAGALADSKD 522
QY 232 TIGSSASVKRESSQSSTQSKEKK-----KKSALDEI-----MEIEEKK--RTA 275

Db 523 EANSMEYDNKSTGRRRRSKGKGNQGLNELKSSSEADDTQVYNRMESQDYKKAPRSK 582

QY 276 RTDYLQPIIIVKIITKK-----LGSKYHKKAIVKEVIDKYTAVVK 317

Db 583 KSTRGMK----TKGTTTKQSSNECDRCGBEFESRTKLHKLADSGHATVK 628

RESULT 13

T34036

hypothetical protein B0041.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004

R:Fulcon, R.; Wohldmann, P.

submitted to the EMBL Data Library, April 1997

A:Description: The sequence of C. elegans cosmid B0041.

A:Reference number: 221466

A:Accession: T34036

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1359 <FUL>

A:Cross-references: UNIPROT:Q9U7B0; UNIPARC:UPI0000060E63; EMBL:AF000196; PIDN:AAC24256.

A:Experimental source: strain Bristol N2; clone B0041

C:Genetics:

A:Gene: CBSP:B0041.7

A:Map position: 1

A:Introns: 12/2; 59/2; 248/2; 582/1; 675/1; 733/3; 966/2; 1044/2; 1310/3

Query Match 6.5%; Score 133; DB 2; Length 1359;

Best Local Similarity 23.7%; Pred. No. 6.4;

Matches 57; Conservative 47; Mismatches 111; Indels 26; Gaps 7;

QY 143 DRDPETIRQLELEKKKKQDLDDDEKTAFTIEQVRRGLEKGEQVPTTTELSRENDEEK 202

Db 131 DSDEEREQSKSKSKTKQTSSSESESEERKVKKKKKEKSVKKGAAETSESEDE 190

QY 203 VTFNLKGCSSSGATSSKSTLGPALKITIGSSASVKESQSQTQSEKKK----- 256

Db 191 KPSKSKGLKKKAKSESESEDEKVKSKKKKSVKKESESEDEAPEKKTKERKR 250

QY 257 -----KKSALDEIMETEIEEKKRTARTDYMLQPIIIVKIITKKLGEYHKKKAIVKEVI-- 309

Db 251 SKTSSSESESEKSDDEEEKESP-----KPKKKPLAVKKLLSDESEESDV-EVL PQ 304

QY 310 -DKYTAVVKMIDSGD----KLKLDQTHL-ETVTPAGPKRILVLNGYRGNEGTLESINEK 363

Db 305 KKKRGAVTLISDEDEKQKSEASDVBEKVKKKAKKQSESESGSDSSEGS1-TVNRK 363

QY 364 T 364

Db 364 S 364

RESULT 14

T21861

hypothetical protein F36F2.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T21861

R:Cottage, A.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19479

A:Accession: T21861

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1877 <WIL>

A:Cross-references: UNIPROT:O62235; UNIPARC:UPI000017B98A; EMBL:Z81532; PIDN:CAB04326.1;

A:Experimental Source: clone F36F2

C:Genetics:

A:Gene: CBSP:F36F2.3

A:Map position: 1

A:Introns: 49/3; 86/2; 112/3; 139/2; 235/3; 284/3; 436/3; 507/2; 566/2; 678/2; 1000/3; 1

Query Match 6.5%; Score 131.5; DB 2; Length 1877;

Best Local Similarity 24.2%; Pred. No. 11;

Matches 72; Conservative 43; Mismatches 113; Indels 69; Gaps 11;

QY 143 DRDPETIRQLELEKKKKQDLDDDEKTAFTIEQVRRGLEKGEQVPTTTELSRENDEEK 202

Db 1020 ERDRERRKDDYEKKSKKESDRONEKEKQREKEVEKEHEKDRKRIKIVKESEKPRKS 1079

QY 203 VTFNLKGCSSSGATSSKSTLGPALK-----TIGSSASVK-----RKSSQSST---- 249

Db 1080 VHERMQK---ADSTSSSSRTTAPSLERKPVSFVASSKPTTNIRVQYSSSSSTKQOE 1136

QY 250 ---QSKEKKKKKKSALDEIMEIEEKKRTARTDYMLQPIIIVKIITKKLGEKY-----HKK 301

Db 1137 DEERSKRDRRKADETDVESIGEKEKKSSRK----VPKESVDVVKHKSFKIFEDLLDFDK 1192

QY 302 KAIVKEVID-----KYTAVVKMIDSGDKLKDQTHLETVI-----PAGV--- 340

Db 1193 SAEINLLMDIKFEGNDAAECTAGLKKASGSLFKDHLIEFKINGKVAGKIKTPCEGVVT 1252

QY 341 -----KRLVLNGYRGNEGTLESINEKTFSATI-----VIETGPLKGRVE 382

Db 1253 FQKGLKPGIVLNKGQ-----VIATVSECTHAIVKDMCATCGKDLREKGRAGORKE 1304

RESULT 15

T28676

rhostry protein - Plasmodium yoelii (fragment)

C:Species: Plasmodium yoelii

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T28676; A45521

R:Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.

Mol. Biochem. Parasitol. 76, 329-332, 1996

A:Title: Comparison of two members of a multigene family coding for high-molecular mass 1

A:Reference number: Z20507; MUID:97077455; PMID:8920022

A:Accession: T28676

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2401 <SIN>

A:Cross-references: UNIPROT:Q26216; UNIPARC:UPI000017B647; EMBL:U36927; NID:g1041784; Pfl

R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.

Mol. Biochem. Parasitol. 42, 241-246, 1990

A:Title: Identification of the gene for a Plasmodium yoelii rhostry protein. Multiple cof

A:Reference number: A45521; MUID:91101660; PMID:2270106

A:Accession: A45521

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 2260-2401 <KEE>

A:Cross-references: UNIPARC:UPI000017B648; GB:M34281

Query Match 6.4%; Score 129.5; DB 2; Length 2401;

Best Local Similarity 21.5%; Pred. No. 19;

Matches 92; Conservative 66; Mismatches 154; Indels 115; Gaps 16;

QY 56 LASENPQQFMDYFSEE-----FRNDF-----FRNDF-----LKL 79

Db 269 LAKEPPYQTKTKIEKNEIYNTIKSDFKIYVGDIEQLYNEFVSQVSNIEHKYTL 328

QY 80 LRRRFGTKRVHNNIVYNE---YISHREHIMNATQWETLDTFTKLGRB----- 125

Db 329 LTLTKIDNVNNIQNMETETVKSHLKNLETNKLSLSETILDIKVIYGEITNELNKLTD 388

QY 126 -----GLC-KVDETPK-----GWYQIVDRDPETIRQLEL-----EKKKKQDLDEEKT 169

Db 389 FKNKEKGLSNKIDYAKENVQLNVYKSNILEIKKHYNQINIDNKEKEAKQNYQDFKSH 448

QY 170 AKFIE-----EQVRRGLEKGEQVPTTTELSRENDEBEKVTFLNSKGCSSSGATSSKSTL 225

Db 449 MKTIPNEMKYQKPSIEIKMKDEFLSKVKNKINDFKVYKEVSEHKNFTLTNK---- 504

QY 226 GPSALKTIGSSASVKRKESSQSSTQSKKKKKKKSALDEIMEIEEKKRTARTDYMLQPI 285

Db 505 ----INTEVSDIEIKYENKFNDKSLINETKKSIEEYQNINTLKK-----VDDY-----I 552

Qy 286 IVKIIITKLGEXHKKKAIVKEVIDXYTAVVMIDSGDKLKD-----OTHLETVI 336
Db 553 KVCINTNELITNCHNKQTTUKDKLNONIKTIKETNSIDKIYTDKPENILTDKKELETKF 612
Qy 337 -----PAPGKRILVNGGYRGNEG-TLESINEKTFSAIIVIETGPLKGREVEGIOYE 387
Db 613 TGLSLNNHESNNKELLTYFYDLCANLGNKENMLYKQFNE-----KEKAVEDIKKK 663
Qy 388 --DISKL 392
Db 664 NVDINKI 670

Search completed: November 23, 2005, 16:34:30
Job time : 27.0638 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

QM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 12:53:47 ; Search time 111.616 Seconds
(without alignments)
10694.773 Million cell updates/sec

Title: US-09-555-529-27

Perfect score: 21

Sequence: 1 tcaaaagacaactgttctggc 21

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_in.*

3: gb_env.*

4: gb_on.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pr.*

9: gb_ro.*

10: gb_ats.*

11: gb_ey.*

12: gb_un.*

13: gb_vi.*

14: gb_intg.*

15: gb_pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	21	100.0	21	6	AX003334 Sequence
2	21	100.0	1102	6	AX003309 Sequence
3	21	100.0	1386	9	BC028860 Mus muscu
4	21	100.0	1390	6	AX003331 Sequence
5	21	100.0	1414	9	XMKN17
6	21	100.0	2710	9	BC058169 Mus muscu
7	21	100.0	180882	14	AC114615 Mus muscu
8	21	100.0	182400	14	AC124010 Mus muscu
9	21	100.0	185378	9	AL772367 Mouse DNA
10	21	100.0	221547	14	AC119716 Rattus no
11	19.4	92.4	144118	14	AC119716 Rattus no
12	19.4	92.4	146360	8	HS1156N12 Human DNA
13	19.4	92.4	153215	14	HS24AI7 Human DNA
14	18.4	87.6	3606	8	BC043228 Homo sapi
15	18.4	87.6	24190	15	CNS09SC4 Oryza sat
16	18.4	87.6	43261	14	AC145382 Oryza sat
17	18.4	87.6	94252	15	AC137616 Oryza sat
18	18.4	87.6	101218	15	AC118284 Oryza sat

C 19	18.4	87.6	108295	15	AP005503	AP005503 Oryza sat
C 20	18.4	87.6	110000	15	AP008214_071	Continuation (72 o
C 21	18.4	87.6	110000	15	AP008214_072	Continuation (73 o
C 22	18.4	87.6	110000	15	AP008214_111	Continuation (112
C 23	18.4	87.6	110000	15	AP008215_007	Continuation (8 of
C 24	18.4	87.6	110000	15	AP008215_087	Continuation (88 o
C 25	18.4	87.6	110000	15	AP008217_081	Continuation (82 o
C 26	18.4	87.6	110000	15	AP008217_121	Continuation (122
C 27	18.4	87.6	110000	15	AP008217_253	Continuation (254
C 28	18.4	87.6	110000	15	AP008218_121	Continuation (122
C 29	18.4	87.6	110000	15	AP008218_122	Continuation (123
C 30	18.4	87.6	110000	15	AP008218_162	Continuation (163
C 31	18.4	87.6	110000	15	AP008218_203	Continuation (204
C 32	18.4	87.6	110000	15	AP008207_194	Continuation (195
C 33	18.4	87.6	110000	15	AP008207_338	Continuation (339
C 34	18.4	87.6	110000	15	AP008208_107	Continuation (108
C 35	18.4	87.6	110000	15	AP008209_084	Continuation (85 o
C 36	18.4	87.6	110000	15	AP008209_174	Continuation (175
C 37	18.4	87.6	110000	15	AP008210_207	Continuation (208
C 38	18.4	87.6	110000	15	AP008210_224	Continuation (225
C 39	18.4	87.6	110000	15	AP008210_321	Continuation (322
C 40	18.4	87.6	110000	15	AP008210_329	Continuation (330
C 41	18.4	87.6	110000	15	AP008211_022	Continuation (23 o
C 42	18.4	87.6	110000	15	AP008211_094	Continuation (95 o
C 43	18.4	87.6	110000	15	AP008211_095	Continuation (96 o
C 44	18.4	87.6	110000	15	AP008211_111	Continuation (112
C 45	18.4	87.6	110000	15	AP008211_216	Continuation (217

ALIGNMENTS

RESULT 1	AX003334	Sequence 27 from Patent WO9929845.	21 bp	DNA	linear	PAT 24-AUG-2000
LOCUS	AX003334					
DEFINITION	AX003334					
ACCESSION	AX003334					
VERSION	AX003334.1	GI:9927147				

SOURCE Mus sp.

ORGANISM Mus sp.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidae; Muridae; Murinae; Mus.

REFERENCE 1 Angulo-Mora,J.F. and Mauffrey,P.

AUTHORS Sequences coding for kin17 protein and their applications

TITLE Patent: WO 9929845-A 27 17-JUN-1999;

JOURNAL ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE

(FR) Location/Qualifiers

FEATURES source

1..21

/organism="Mus sp."

/mol_type="unassigned DNA"

/db_xref="taxon:10095"

ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 21;

Best Local Similarity 100.0%; Pred. No. 9.6;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCAAAGACAACACTGTGTCTGGC 21

Db 1 TCAAAGACAACACTGTGTCTGGC 21

RESULT 2

AX003309

LOCUS

DEFINITION Sequence 2 from Patent WO9929845.

ACCESSION AX003309

VERSION AX003309.1 GI:9927126

KEYWORDS

linear

DNA

1102 bp

PAT 24-AUG-2000

SOURCE Mus sp.
ORGANISM Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.

REFERENCE 1
AUTHORS Angulo-Mora, J.F. and Mauffrey, P.
TITLE Sequences coding for kin17 protein and their applications
JOURNAL Patent: WO 9929845-A 2 17-JUN-1999;
(FR) ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE (FR)

FEATURES
source Location/Qualifiers
1..1102 /organism="Mus sp."
/mol_type="unassigned DNA"
/db_xref="taxon:10095"

ORIGIN
Query Match 100.0%; Score 21; DB 6; Length 1102;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAAGACACAACTGTTGCTGGC 21
|||||
Db 150 TCAAGACACAACTGTTGCTGGC 170

RESULT 3
LOCUS BC028860
DEFINITION Mus musculus antigenic determinant of rec-A protein, mRNA (CDNA clone MGC:18413 IMAGE:3673830), complete cds.
ACCESSION BC028860
VERSION BC028860.1 GI:22135637
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1386)
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Schetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, E., Kettaman, M., Madan, A., Rodriguez, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalls, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED 12477932
REFERENCE 2 (bases 1 to 1386)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (01-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgabbs-remail.nih.gov

Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louleaged, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 23 Row: h Column: 8
This clone was selected for full length sequencing because it passed the following selection criteria: GenomesScan gene prediction.

FEATURES
source Location/Qualifiers
1..1386 /organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:18413 IMAGE:3673830"
/tissue_type="Mammary tumor. C3(1)-Tag model. Infiltrating ductal carcinoma. 5 month old virgin mouse."
/clone_lib="NCI CGAP_Mam6"
/lab_host="DH10B"
/note="Vector: PCMV-SPORT6"
1..1386 /gene="Kin"
/note="Synonym: Kin17"
/db_xref="GeneID:16588"
/db_xref="MGI:96676"
20..1195 /gene="Kin"
/codon_start=1
/product="antigenic determinant of rec-A protein"
/protein_id="AAH28860.1"
/db_xref="GI:22135638"
/db_xref="GeneID:16588"
/db_xref="MGI:96676"

gene
CDS
Query Match 100.0%; Score 21; DB 9; Length 1386;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAAGACACAACTGTTGCTGGC 21
|||||
Db 169 TCAAGACACAACTGTTGCTGGC 189

ORIGIN
Query Match 100.0%; Score 21; DB 9; Length 1386;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 4
LOCUS AX003331
DEFINITION Sequence 24 from Patent WO9929845.
ACCESSION AX003331
VERSION AX003331.1 GI:9927146
KEYWORDS Mus sp.
SOURCE Mus sp.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.

```

REFERENCE
AUTHORS      1 Angulo-Mora,J.F. and Mauffrey,P.
TITLE        Sequences coding for kin17 protein and their applications
JOURNAL      Patent: WO 9929845-A 24 17-JUN-1999;
              ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE
              (FR)
FEATURES
source       1. 1390
              /location/Qualifiers
              /organism="Mus sp."
              /mol_type="unassigned DNA"
              /db_xref="taxon:10095"
ORIGIN
Query Match      100.0%; Score 21; DB 6; Length 1390;
Best Local Similarity 100.0%; Pred. NO. 8.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TCAAAGACAACTGTTGCTGGC 21
        |||||||
Db      150 TCAAAGACAACTGTTGCTGGC 170

RESULT 5
MKKIN17
LOCUS      1414 bp mRNA linear ROD 30-OCT-1995
DEFINITION Mouse KIN17 mRNA for kin17 protein.
ACCESSION X58472
VERSION   X58472.1 GI:1045209
KEYWORDS  KIN17 gene; kin17 protein; zinc-finger protein.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE
AUTHORS    1 (bases 1 to 1414)
            Angulo,J.F., Rouer,E., Benarous,R. and Devoret,R.
TITLE      Identification of a mouse cDNA fragment whose expressed polypeptide
            reacts with anti-recA antibodies
JOURNAL    Biochimie 73 (2-3), 251-256 (1991)
PUBMED    1715759
REFERENCE
AUTHORS    2
            Angulo,J.F., Rouer,E., Mazin,A., Mattei,M.G., Tissier,A.,
            Horellou,P., Benarous,R. and Devoret,R.
TITLE      Identification and expression of the cDNA of KIN17, a zinc-finger
            gene located on mouse chromosome 2, encoding a new DNA-binding
            protein
JOURNAL    Nucleic Acids Res. 19 (19), 5117-5123 (1991)
PUBMED    1923796
REFERENCE
AUTHORS    3
            Angulo,J.
TITLE      Direct Submission
JOURNAL    Submitted (25-MAR-1991) J. Angulo, C N R S, GEMC-Lab d'Enzymologie,
            1 Avenue de la Terrasse, Gif-Sur-Yvette, Ile-de-France, FRANCE
REMARK     revised by [4] MGP
REFERENCE
AUTHORS    4 (bases 1 to 1414)
            Angulo,J.
TITLE      Direct Submission
JOURNAL    Submitted (30-OCT-1995) J. Angulo, C N R S, GEMC-Lab d'Enzymologie,
            1 Avenue de la Terrasse, Gif-Sur-Yvette, Ile-de-France, FRANCE
COMMENT    On Oct 31, 1995 this sequence version replaced gi:52795.
FEATURES
source     1. 1414
            /location/Qualifiers
            /organism="Mus musculus"
            /mol_type="mRNA"
            /db_xref="taxon:10090"
            /cell_line="MBE6"
            /tissue_type="Brain"
            /dev_stage="embryo"
            1. 1414
            /gene="KIN17"
            /pseudo
            <1..1382
            /gene="KIN17"

```

```

CDS
/pseudo
25..1200
/gene="KIN17"
/pseudo
/codon_start=1
/product="kin17 protein"
/db_xref="PSEUDO:CAA41386.1"
106..174
/gene="KIN17"
/notes="zinc-finger of kin17"
/pseudo
234..842
/gene="KIN17"
/notes="kin17(200) fragment"
/pseudo
/evidence=experimental
472..496
/gene="KIN17"
/notes="kin17 helix1 motif"
/pseudo
781..792
/gene="KIN17"
/notes="nuclear localization signal 1"
/pseudo
907..924
/gene="KIN17"
/notes="nuclear localization signal 2"
/pseudo
1382
/gene="KIN17"
polyA_site
ORIGIN
Query Match      100.0%; Score 21; DB 9; Length 1414;
Best Local Similarity 100.0%; Pred. NO. 8.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TCAAAGACAACTGTTGCTGGC 21
        |||||||
Db      174 TCAAAGACAACTGTTGCTGGC 194

RESULT 6
BC058169
LOCUS      2710 bp mRNA linear ROD 13-FEB-2004
DEFINITION Mus musculus antigenic determinant of rec-A protein, mRNA (cDNA
            clone IMAGE:5026202), partial cds.
ACCESSION BC058169
VERSION   BC058169
KEYWORDS  Mus musculus (house mouse)
SOURCE    Mus musculus
ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE
AUTHORS    1 (bases 1 to 2710)
            Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Ustin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Kettaman,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Green,E.D.,
            Bouffard,G.G., Blakesley,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

```

TITLE
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
REFERENCE
12477932
AUTHORS
Strausberg, R.
TITLE
Direct Submission
JOURNAL
Submitted (15-SEP-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT
Contact: MGC help desk
Email: cgapsb@mail.nih.gov
Tissue Procurement: Lohar Hennighausen Ph.D., Robin Humphreys cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland
Web site: <http://www.nisc.nih.gov/>
Contact: nisc.mgc@hgi.nih.gov
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Lalic, P., Legaspi, R., Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Tsurgueon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 123 Row: d Column: 16
This clone was selected for full length sequencing because it passed the following selection criteria: Genomescan gene prediction, Similarity but not identity to protein.
FEATURES
Location/Qualifiers
1..2710
/organism="Mus musculus"
/mol_type="mRNA"
/strain="mix FVB/N, C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:5026202"
/tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months old, gross tissue."
/clone_lib="NCI CGAP Mam5"
/lab_host="DH105"
/note="Vector: pCMV-SPORT6"
ORIGIN
Query Match 100.0%; Score 21; DB 9; Length 2710;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY
1 TCAAGACAACTGTTGCTGCG 21
|||||
198 TCAAGACAACTGTTGCTGCG 218
Db
AC114615 180882 bp DNA linear HTG 09-MAR-2004
Mus musculus clone RP24-82119, WORKING DRAFT SEQUENCE, 4 ordered pieces.
ACCESSION
AC114615
VERSION
AC114615.5 GI:44886278
KEYWORDS
HTG; HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidae; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

1 (bases 1 to 180882)
Birren, B., Nusbaum, C. and Lander, E.
Mus musculus, Clone RP24-82119
Unpublished
2 (bases 1 to 180882)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazarar, B., Chopel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fato, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Coyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karatas, A., Kelle, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (10-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 180882)
Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArrellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kelle, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (09-MAR-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 2, 2004 this sequence version replaced gi:24182248.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L24315
Center clone name: 82_1_19
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 191018 bases at least Q40
Consensus quality: 191303 bases at least Q30
Consensus quality: 191388 bases at least Q20

TITLE
JOURNAL
COMMENT


```

Insert size: 178000; agarose-fp
Insert size: 191574; sum-of-contigs
Quality coverage: 12.2 in Q20 bases; agarose-fp
Quality coverage: 11.3 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 89963: contig of 89963 bp in length
* 89964 90063: gap of 100 bp
* 90064 94560: contig of 4497 bp in length
* 94561 94660: gap of 100 bp
* 94661 169601: contig of 74941 bp in length
* 169602 169701: gap of 100 bp
* 169702 180882: contig of 11181 bp in length.

FEATURES
    source
    1. 180882
       /organism="Mus musculus"
       /mol_type="genomic DNA"
       /db_xref="taxon:10090"
       /clone="RP24-82119"
       /clone_lib="RPC1-24 Male Mouse BAC"
    misc_feature
    1. 89963
       /note="assembly_fragment"
       clone_end:SP6
       vector_side:left
    gap
    89964..90063
       /estimated_length=100
    misc_feature
    90064..94560
       /note="assembly_fragment"
    gap
    94561..94660
       /estimated_length=100
    misc_feature
    94661..169601
       /note="assembly_fragment"
    gap
    169602..169701
       /estimated_length=100
    misc_feature
    169702..180882
       /note="assembly_fragment"
       clone_end:T7
       vector_side:right

ORIGIN
Query Match      100.0%; Score 21; DB 14; Length 180882;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCAAAGACAACTGTTGCTGGC 21
    |||||
Db 52817 TCAAAGACAACTGTTGCTGGC 52797

RESULT 8
AC124010
LOCUS
DEFINITION Mus musculus chromosome UNK clone RP23-119N4, WORKING DRAFT
AC124010 182400 bp DNA linear HTG 25-AUG-2002
SEQUENCE 6 unordered pieces.
AC124010
AC124010.2 Gi:22475963
VERSION
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 182400)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE The sequence of Mus musculus clone

```

```

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 182400)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (06-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 182400)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT On Aug 25, 2002 this sequence version replaced gi:21327639.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M_BA0119N04
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 183528 bases at least Q40
Consensus quality: 184210 bases at least Q30
Consensus quality: 184756 bases at least Q20
Insert size: 197000; agarose-fp
Quality coverage: 9.53 in Q20 bases; agarose-fp
Quality coverage: 9.61 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 11868: contig of 11868 bp in length
* 11869 11968: gap of unknown length
* 11969 28906: contig of 16938 bp in length
* 28907 29006: gap of unknown length
* 29007 72773: contig of 43767 bp in length
* 72774 72873: gap of unknown length
* 72874 124926: contig of 52053 bp in length
* 124927 125026: gap of unknown length
* 125027 181902: contig of 56876 bp in length
* 181903 182002: gap of unknown length
* 182003 182400: contig of 398 bp in length.

FEATURES
    source
    1. 182400
       /organism="Mus musculus"
       /mol_type="genomic DNA"
       /db_xref="taxon:10090"
       /chromosome="UNK"
       /clones="RP23-119N4"
    misc_feature
    1. 11868
       /note="assembly_name:Contig86"
    gap
    11869..11968
       /estimated_length=unknown
    misc_feature
    11969..28906
       /note="assembly_name:Contig87"
    gap
    28907..29006
       /estimated_length=unknown
    misc_feature
    29007..72773
       /note="assembly_name:Contig88"
    gap
    72774..72873
       /estimated_length=unknown

```

```

misc_feature 72874..124926
              /note="assembly_name:Contig89"
gap          124927..125026
              /estimated_length=unknown
misc_feature 125027..181902
              /note="assembly_name:Contig90"
gap          181903..182002
              /estimated_length=unknown
misc_feature 182003..182400
              /note="assembly_name:Contig16"

ORIGIN
Query Match      100.0%; Score 21; DB 14; Length 182400;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGGC 21
    |||||
Db 162596 TCAAAGACAACTGTTGCTGGC 162616

RESULT 9
AL772367/c      185378 bp DNA linear ROD 02-SBP-2002
LOCUS           Mouse DNA sequence from clone RP23-119N4 on chromosome 2, complete
DEFINITION
ACCESSION       AL772367
VERSION         AL772367.7 GI:22759512
KEYWORDS        HTG.
SOURCE          Mus musculus (house mouse)
ORGANISM        Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE
AUTHORS         Wood,J.
TITLE           Direct Submission
JOURNAL         Submitted (30-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
                Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
                humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
                On Sep 9, 2002 this sequence version replaced gi:2265488.
COMMENT         ----- Genome Center
                Center: Wellcome Trust Sanger Institute
                Center code: SC
                Web site: http://www.sanger.ac.uk
                Contact: humquery@sanger.ac.uk
                -----
                During sequence assembly data is compared from overlapping clones.
                Where differences are found these are annotated as variations
                together with a note of the overlapping clone name. Note that the
                variation annotation may not be found in the sequence submission
                corresponding to the overlapping clone, as we submit sequences with
                only a small overlap as described above.
                This sequence was finished as follows unless otherwise noted: all
                regions were either double-stranded or sequenced with an alternate
                chemistry or covered by high quality data (i.e., phred quality >=
                30); an attempt was made to resolve all sequencing problems, such
                as compressions and repeats; all regions were covered by at least
                one plasmid subclone or more than one M13 subclone; and the
                assembly was confirmed by restriction digest. The following
                abbreviations are used to associate primary accession numbers given
                in the feature table with their source databases: Em:, EMBL; Sw:,
                SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
                database can be found at
                http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-119N4 is
                from the RP23-Mouse PAC library
                constructed by the group of Pieter de Jong.
                For further details see http://www.chori.org/bacpac/home.htm
                VECTOR: pBACe3.6.
FEATURES
source          1..185378
                /location=Qualifiers
                /organism="Mus musculus"
                /mol_type="genomic DNA"

```

```

/db_xref="taxon:10090"
/Chromosomes="2"
/clone_lib="RPC1-23"

ORIGIN
Query Match      100.0%; Score 21; DB 9; Length 185378;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGGC 21
    |||||
Db 146231 TCAAAGACAACTGTTGCTGGC 146211

RESULT 10
AL119716
LOCUS           Rattus norvegicus clone CH230-274116, *** SEQUENCING IN PROGRESS
DEFINITION
ACCESSION       AC119716
VERSION         AC119716.6 GI:25092366
KEYWORDS        HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE          Rattus norvegicus (Norway rat)
ORGANISM        Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidae; Muridae; Murinae; Rattus.
REFERENCE
AUTHORS         Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
                Allen,C., Allen,H., Alebrooks,S., Amin,A., Anguiano,D.,
                Anyalebechi,V., Ayodeji,A., Ayodeji,M., Baca,E., Baden,H.,
                Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
                Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
                Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
                Cardenas,V., Carter,K., Cavazos,I., Ceasear,H., Center,A.,
                Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
                Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
                Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
                Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
                Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
                Egan,A., Ecatotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
                Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
                Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
                Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
                Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
                Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
                Hollings,B., Howells,S., Hladun,S.L., Hodgson,A., Hogues,M.,
                Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
                Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
                Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
                Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
                Lorenshewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
                Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
                Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
                Mathew,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
                Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
                Morgan,M., Morris,K., Morris,S., Muniadasa,M., Murphy,M., Naik,L.,
                Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
                Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
                Pasceriak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
                Plopper,F., Poindexter,A., Popovic,D., Primus,E., Fu,L.-L.,
                Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
                Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
                Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
                Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
                Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajis,D.,
                Speed,A., Sodergren,B., Song,X.-Z., Sorelle,R., Sosa,J.,
                Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
                Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Uman,K.,
                Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
                Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,P.,

```

Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausern, A., Weises, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weinstock, G., and Gibbs, R.A.
 Direct Submission
 Unpublished
 2 (bases 1 to 221547)
 Worley, K.C.
 Direct Submission
 Submitted (01-MAY-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 221547)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Nov 19, 2002 this sequence version replaced gi:23611286.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GWLC
 Center clone name: CH230-274116
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 208013 bases at least Q40
 Consensus quality: 211685 bases at least Q30
 Consensus quality: 213907 bases at least Q20
 Estimated insert size: 199182; sum-of-contigs estimation
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 17 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 198201: contig of 198201 bp in length
 * 198202: 198301: gap of unknown length
 * 198302 199749: contig of 1448 bp in length
 * 199750 199849: gap of unknown length
 * 199850 201115: contig of 1266 bp in length
 * 201116 201215: gap of unknown length
 * 201216 202338: contig of 1123 bp in length
 * 202339 202439: gap of unknown length
 * 202439 203500: contig of 1062 bp in length
 * 203501 203601: gap of unknown length
 * 203601 204905: contig of 1305 bp in length
 * 204906 205005: gap of unknown length
 * 205006 206192: contig of 1187 bp in length
 * 206193 206293: gap of unknown length
 * 206293 207385: contig of 1053 bp in length

* 207386 207485: gap of unknown length
 * 207486 208519: contig of 1034 bp in length
 * 208520 209758: gap of unknown length
 * 209759 209858: gap of unknown length
 * 209859 211441: contig of 1583 bp in length
 * 211442 211542: gap of unknown length
 * 211543 212876: contig of 1335 bp in length
 * 212877 212976: gap of unknown length
 * 212977 214256: contig of 1280 bp in length
 * 214257 214357: gap of unknown length
 * 214358 217003: contig of 2647 bp in length
 * 217004 217104: gap of unknown length
 * 217105 218459: contig of 1356 bp in length
 * 218460 218559: gap of unknown length
 * 218560 219788: contig of 1229 bp in length
 * 219789 219888: gap of unknown length
 * 219889 221547: contig of 1659 bp in length.

FEATURES
 source
 1. .221547
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-274116"
 misc_feature
 1. .1272
 /note="wgs end extension
 clone_end:Sp6"
 /note="clone_boundary
 clone_end:Sp6"
 site:
 end sequence:BZ266352"
 192445..193553
 /note="wgs contig"
 195743..196592
 /note="clone_boundary
 clone_end:T7"
 site:
 end sequence:BZ266349"
 198202..198301
 /estimated_length=unknown
 199750..199849
 /estimated_length=unknown
 201116..201215
 /estimated_length=unknown
 202339..202438
 /estimated_length=unknown
 203501..203600
 /estimated_length=unknown
 204906..205005
 /estimated_length=unknown
 206193..206292
 /estimated_length=unknown
 207386..207485
 /estimated_length=unknown
 208520..208619
 /estimated_length=unknown
 209759..209858
 /estimated_length=unknown
 211442..211541
 /estimated_length=unknown
 212877..212976
 /estimated_length=unknown
 214257..214356
 /estimated_length=unknown
 217004..217103
 /estimated_length=unknown
 218460..218559
 /estimated_length=unknown
 219789..219888
 /estimated_length=unknown
 Query Match 100.0%; Score 21; DB 14; Length 221547;

Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAAGACAACTGTTGCTGCG 21
|||||

Db 213161 TCAAGACAACTGTTGCTGCG 213181

RESULT 11
AC141995/c
LOCUS
DEFINITION AC141995 144118 bp DNA linear HTG 21-MAR-2003
*** 48 unordered pieces.

ACCESSION AC141995
VERSION AC141995.1 GI:29135466
KEYWORDS HTG; HTGS PHASE1.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Murioidea; Muridae; Murinae; Rattus;
1 (bases 1 to 144118)

REFERENCE
Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alebrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganti, R., Garcia, A., Garner, T., Garza, M.,
Georgiev, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensu, L., Louleghed, H., Lozado, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhiney, S., Mcleod, M., McNeill, T., Meenen, E., Milosavljevic, A.,
Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K.,
Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D.,
Newton, N., Nguyen, N., Norris, S., Nwaokeme, O., Okwuonu, G.,
Olarunpuga, A., Pal, S., Parks, K., Pasternak, S., Paul, H.,
Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A.,
Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E.,
Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y.,
Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A.,
Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S.,
Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A.,
Sisson, I., Sitter, C. D., Smajda, D., Sneed, A., Sodergren, E.,
Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A.,
Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S.,
Tingey, A., Trejos, Z., Umani, K., Valas, R., Vera, V., Villaseana, D.,
Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J.,
Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R.,
Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S.,
Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X.,
Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R.,
Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Direct Submission
Unpublished
2 (bases 1 to 144118)

AUTHORS TITLE JOURNAL

COMMENT

Worley, K. C.
Direct Submission
Submitted (21-MAR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GRCN
Center Clone name: CH230-293H7
----- Summary Statistics
Sequencing vector: Plasmid,
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 108995 bases at least Q40
Consensus quality: 114645 bases at least Q30
Consensus quality: 119503 bases at least Q20
Estimated insert size: 114401; sum-of-contigs estimation
Quality coverage: 2x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 48 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1394: contig of 1394 bp in length
* 1395 1494: gap of unknown length
* 1495 2950: contig of 1356 bp in length
* 2851 2950: gap of unknown length
* 2951 3372: contig of 1022 bp in length
* 3372 4072: gap of unknown length
* 4073 5309: contig of 1237 bp in length
* 5310 5409: gap of unknown length
* 5410 6564: contig of 1155 bp in length
* 6565 6664: gap of unknown length
* 6665 8449: contig of 1785 bp in length
* 8450 8449: gap of unknown length
* 8550 9904: contig of 1354 bp in length
* 9904 10003: gap of unknown length
* 10004 11746: contig of 1743 bp in length
* 11747 11846: gap of unknown length
* 11847 13518: contig of 1672 bp in length
* 13519 13618: gap of unknown length
* 13619 14970: contig of 1352 bp in length
* 14971 15070: gap of unknown length
* 15071 16373: contig of 1303 bp in length
* 16374 16473: gap of unknown length
* 16474 18359: contig of 1886 bp in length
* 18360 18459: gap of unknown length
* 18460 19821: contig of 1362 bp in length
* 19822 19921: gap of unknown length
* 19922 21449: contig of 1528 bp in length
* 21450 21549: gap of unknown length
* 21550 23269: contig of 1720 bp in length
* 23270 23369: gap of unknown length
* 23370 25710: contig of 2341 bp in length
* 25711 25810: gap of unknown length
* 25811 27760: contig of 1950 bp in length
* 27761 27860: gap of unknown length
* 27861 29815: contig of 1955 bp in length
* 29816 29915: gap of unknown length
* 29916 31213: contig of 1298 bp in length
* 31214 31313: gap of unknown length
* 31314 33042: contig of 1729 bp in length
* 33043 33142: gap of unknown length

TITLE JOURNAL REFERENCE

* 33143	34194: contig of 1052 bp in length	HS1156N12	146360 bp	DNA	linear	PRI 18-MAY-2005
* 34195	34294: gap of unknown length	LOCUS				
* 34295	36020: contig of 1726 bp in length	DEFINITION				
* 36021	36120: gap of unknown length	ACCESSION	AL009047			
* 36121	37671: contig of 1551 bp in length	VERSION	AL009047.1	GI:3355872		
* 36122	37771: gap of unknown length	KEYWORDS	HTG			
* 37672	37759: contig of 1988 bp in length	SOURCE	Homo sapiens (human)			
* 37772	39859: gap of unknown length	ORGANISM	Homo sapiens			
* 37760	42442: contig of 2583 bp in length					
* 39860	42542: gap of unknown length					
* 42443	44482: contig of 1940 bp in length					
* 42543	44582: gap of unknown length					
* 44483	46351: contig of 1769 bp in length					
* 44583	46451: gap of unknown length					
* 46352	48712: contig of 2261 bp in length	REFERENCE	1 (bases 1 to 146360)			
* 46452	48812: gap of unknown length	AUTHORS	Bird,C.			
* 48713	51252: contig of 2440 bp in length	TITLE	Direct Submission			
* 51253	51352: gap of unknown length	JOURNAL	Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk			
* 51353	54232: contig of 2880 bp in length		Clone requests: clonesrequest@sanger.ac.uk			
* 54233	54332: gap of unknown length		On Jul 30, 1998 this sequence version replaced gi:3334529.			
* 54333	57136: contig of 2804 bp in length	COMMENT	The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:			
* 57137	57236: gap of unknown length		Em., EMBL; Swi., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at			
* 57237	59833: contig of 2597 bp in length		http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at			
* 59834	62619: contig of 2686 bp in length		http://www.sanger.ac.uk/HGP/ChrX			
* 62620	62719: gap of unknown length		RP5-1156N12 is from the library RPCI-5 constructed by the group of Pieter de Jong. For further details see			
* 62720	65087: contig of 2368 bp in length		http://www.chori.org/bacpac/home.htm			
* 65088	65187: gap of unknown length		VECTOR: pCYPAC2			
* 65188	68329: contig of 3142 bp in length		----- Genome Center			
* 68330	68429: gap of unknown length		Center: Wellcome Trust Sanger Institute			
* 68430	72825: contig of 4396 bp in length		Center code: SC			
* 72826	72925: gap of unknown length		Web site: http://www.sanger.ac.uk			
* 72926	76196: contig of 3271 bp in length		Contact: vegas@sanger.ac.uk			
* 76197	76296: gap of unknown length		-----			
* 76297	80327: contig of 4031 bp in length		This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.			
* 80328	80427: gap of unknown length	FEATURES	Location/Qualifiers			
* 80428	83596: contig of 3169 bp in length	source	1. 146360			
* 83597	83696: gap of unknown length		/organism="Homo sapiens"			
* 83697	88053: contig of 4357 bp in length		/mol_type="genomic DNA"			
* 88054	88153: gap of unknown length		/db_xref="RZPD:RPCIP704N121156"			
* 88154	91623: contig of 3470 bp in length		/db_xref="taxon:9606"			
* 91624	91723: gap of unknown length		/clone="RP5-1156N12"			
* 91724	96501: contig of 4778 bp in length		/clone_lib="RPCI-5"			
* 96502	96601: gap of unknown length	ORIGIN				
* 96602	101193: contig of 4592 bp in length	Query Match	92.4%; Score 19.4; DB 8; Length 146360;			
* 101194	101293: gap of unknown length	Best Local Similarity	95.2%; Pred. No. 48;			
* 101294	106768: contig of 5475 bp in length	Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
* 106769	106868: gap of unknown length					
* 106869	111636: contig of 4768 bp in length	Qy	1 TCAAAGACAACTGTTGCTGGC 21			
* 111637	111736: gap of unknown length	Db	56392 TCAAAGACAACTGATGCTGGC 56412			
* 111737	117842: contig of 6106 bp in length					
* 117843	117942: gap of unknown length	RESULT 13				
* 117943	130224: contig of 12282 bp in length	HS24A17	153215 bp	DNA	linear	HTG 10-JUL-2001
* 130225	130324: gap of unknown length	LOCUS				
* 130325	144118: contig of 13794 bp in length.	DEFINITION	Homo sapiens chromosome X clone RP6-24A17 map q21.1-21.33.			
		ACCESSION	AL035452			
		VERSION	AL035452.7	GI:10045108		
		KEYWORDS	HTG; HTGS_PHASE2; HTGS_CANCELLED.			


```
/db_xref="GI:71297022"
/db_xref="GeneID:339318"
/db_xref="MIM:606741"
/translation="MMEKLSKGMIPDWSRWENKELSTYKKNYDSDSPQTVIIKVV
KQSYFBSNKKLEIYETKLEGGHGSQVDHFRPAITLSRESPTADSVYKYNIPRSTPHS
KSTLSFQPKISAEKNSHKYDILKKNLPKKSIVKNEKVGKLLNNGKGAAPSOQKS
LTLPTCNREKIYTCSECGKAFQKQSLNHRWRIHTGKPYECRECGKTFSHGSSLTR
HLISHGEKPYKIECGKAFSHVSLTNHOSHTGKPYECMNCXGFSRVSHLIEHL
RIHTQKLYECRICGKAFTHRSSLIHHQIHTGKPYECRECGKAFCCSSHLTRHQRI
HTMEKYECNCKLKVFSUSFLVQHOSHTGKPYECRECGKAFCCSSHLTRHQRI
RLKPYECSCGKAFSHRHSIHLQHRHTGKPYECRECGKAFCCSSHLTRHQRIHTE
KPYKQNECGKAFSGKSNLTAHQRVHNGKPNVSVEKPLDYNMHTCKEYSYRRTV"
```

ORIGIN

```
Query Match      87.6%; Score 18.4; DB 8; Length 3606;
Best Local Similarity 95.0%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 CAAAGACAACTGTTGCTGGC 21
      |||||
Db      391 CAAAGACAACTGTTGCTGGC 372
```

RESULT 15

```
CNS09SC4      24190 bp      DNA      linear      PLN 29-JAN-2004
LOCUS      Oryza sativa chromosome 12, . Partial sequence from BAC B1004E12 of
DEFINITION      chromosome 12 of cultivar Nipponbare of ssp. japonica of Oryza
      sativa (rice), complete sequence.
```

```
ACCESSION      BX672566
VERSION      BX672566.2 GI:41393044
```

KEYWORDS

```
SOURCE      HTG.
```

ORGANISM

```
Oryza sativa (japonica cultivar-group)
```

```
Oryza sativa (japonica cultivar-group)
```

```
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```

```
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
```

```
Ehrhartoideae; Oryzaceae; Oryza.
```

```
1 (bases 1 to 24190)
```

```
Choisne, N., Orjeda, G., Cattolico, L., Demange, N., Wincker, P.,
```

```
Segurens, B., Pelletier, E., Scarpelli, C., Salanoubat, M.,
```

```
Weissenbach, J. and Quetier, F.
```

```
Oryza sativa chromosome 12 sequencing
```

```
Unpublished
```

```
2 (bases 1 to 24190)
```

```
Genoscope.
```

```
Direct Submission
```

```
Submitted (29-JAN-2004) Genoscope - Centre National de Sequencage :
```

```
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
```

```
- Web : www.genoscope.cns.fr)
```

```
On Jan 29, 2004 this sequence version replaced gi:37936142.
```

```
Center: Genoscope / Centre National de Sequencage
```

```
Center code: GS
```

```
Web site: http://www.genoscope.cns.fr/
```

```
Contact: SeqRef@genoscope.cns.fr
```

```
-----
```

```
The following sequence is oriented from the T7 to the SP6 end.
```

```
Upstream BAC (overlapping the T7 end) : B1097F04 (AC=BX649219)
```

```
Downstream BAC (overlapping the SP6 end) : OSJNB0119N22
```

```
(AC=AL732380) WARNING: This sequence only represents a part of the
```

```
indicated clone. ----- Finishing boundaries
```

```
FINISHED SEGMENT STARTS AT BASE 1
```

```
FINISHED SEGMENT ENDS AT BASE 24190
```

```
-----
```

FEATURES

```
source
```

```
1. :24190      Location/Qualifiers
```

```
    /organism="Oryza sativa (japonica cultivar-group)"
```

```
    /mol_type="genomic DNA"
```

```
    /cultivar="Nipponbare"
```

```
    /sub_species="japonica"
```

```
    /db_xref="taxon:39947"
```

```
    /chromosome="12"
```

```
    /clone="B1004E12"
```

ORIGIN

```
Query Match      87.6%; Score 18.4; DB 15; Length 24190;
Best Local Similarity 95.0%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 TCAAAGACAACTGTTGCTGG 20
```

```
      |||||
```

```
Db      149 TCAAAGAGAACTGTTGCTGG 168
```

```
Search completed: November 26, 2005, 19:11:11
Job time : 115.616 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 12:48:57 ; Search time 13.4698 Seconds
(without alignments)
10390.557 Million cell updates/sec

Title: US-09-555-529-27

Perfect score: 21

Sequence: 1 tcaagaacaacttctgctggc 21

Scoring table: IDENTITY_NUC

Gapop 10,0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N Geneseq_21.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1990s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001as.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002as.*
- 7: Geneseqn2002bs.*
- 8: Geneseqn2003as.*
- 9: Geneseqn2003bs.*
- 10: Geneseqn2003cs.*
- 11: Geneseqn2003ds.*
- 12: Geneseqn2004as.*
- 13: Geneseqn2004bs.*
- 14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	21	2	AAX85571 PCR prime
2	21	100.0	1102	2	AAX85550 CDNA of a
3	21	100.0	1390	2	AAX85570 CDNA of a
4	21	100.0	1458	2	AAX79936 Murine Ki
5	18.4	87.6	10319	4	AAX27632 DNA encod
6	18.4	87.6	10319	10	ADB94435 Novel hum
7	17.8	84.8	21	2	AAX85568 PCR prime
8	17.8	84.8	1002	2	AAX85552 Probe der
9	17.8	84.8	1002	2	AAX85551 CDNA of a
10	17.8	84.8	1296	2	AAX85549 CDNA of a
11	17.8	84.8	12555	4	ABK42300 Genomic s
12	17.8	84.8	12555	4	ABK42300 Human imm
13	17.8	84.8	12555	9	ADB60456 Connectiv
14	17.8	84.8	13810	4	ABU06020 Drosophil
15	16.8	80.0	329	13	ACF85421 Human SIR
16	16.8	80.0	780	9	ADA29641 DNA encod
17	16.8	80.0	2795	14	ADZ61207 Murine Nu
18	16.8	80.0	56423	10	ADC85728 Human GPC
19	16.8	80.0	80331	4	AAC89559 Human his

20	16.8	80.0	110000	13	ABD32911_3	Continuation (4 of
21	16.8	80.0	203020	14	ADZ12576	Adz12576 Human can
22	16.4	78.1	561	10	ADK63809	Adk63809 Disease t
23	16.4	78.1	561	10	ADK61845	Adk61845 Disease t
24	16.4	78.1	939	4	AAH31803	Aah31803 Human Olf
25	16.4	78.1	951	5	AAH42275	Aah42275 Human CDN
26	16.4	78.1	951	6	ABZ43011	Abz43011 Human GPC
27	16.4	78.1	951	6	ABK68432	Abk68432 Human DNA
28	16.4	78.1	951	6	ABK37561	Abk37561 DNA encod
29	16.4	78.1	951	12	ADG83307	Adg83307 Human Olf
30	16.4	78.1	971	6	AAH40634	Aah40634 Human G-p
31	16.4	78.1	980	6	ABQ88378	Abq88378 Human G-p
32	16.4	78.1	980	6	ABQ88379	Abq88379 Human G-p
33	16.4	78.1	980	10	ADI04292	Adi04292 Human G-p
34	16.4	78.1	980	10	ADI04294	Adi04294 Human G-p
35	16.4	78.1	1351	10	ADC85738	Adc85738 Human GPC
36	16.4	78.1	1885	3	AAC42432	Aac42432 Arabidops
37	16.4	78.1	1959	13	ADT15885	Adt15885 Plant CDN
38	16.4	78.1	2000	8	ADA72658	Ada72658 Rice gene
39	16.4	78.1	110000	14	ABE39174_01	Continuation (2 of
40	16.4	78.1	110000	14	ABE39174_02	Continuation (3 of
41	16.4	78.1	110000	14	ABE39175_17	Continuation (18 o
42	16.4	78.1	110000	14	ABE39175_18	Continuation (19 o
43	16.4	78.1	110000	14	ABE42401_15	Continuation (16 o
44	16.4	78.1	110000	14	ABE42737_04	Continuation (5 of
45	16.4	78.1	110000	14	ABE35723_1	Continuation (2 of

ALIGNMENTS

RESULT_1

AAX85571
ID AAX85571 standard; cDNA; 21 BP.

XX AAX85571;

DT 07-SEP-1999 (first entry)

DE PCR primer used to amplify human kin17 cDNA sequences.

XX Human; kin17 protein; cell proliferation; fertility;

KW hyperproliferative disease; protein interaction; curved DNA;

KW HIV replication; HIV integration; repair enzyme; PCR primer; ss.

XX Synthetic.

OS Mus sp.

PN FR2772046-A1.

XX 11-JUN-1999.

XX 09-DEC-1997; 97FR-00015536.

XX 09-DEC-1997; 97FR-00015536.

XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;

XX WPI; 1999-359999/31.

XX New DNA coding for human kin17 protein - useful for controlling cell

PT proliferation or fertility.

XX Example 1; Page 12; 69pp; French.

XX PCR primers AAX85571-72 were used to amplify 1000 base pairs of cDNA
encoding a human kin17 protein. The mammalian kin17 protein is useful for
preparing a medicament for controlling cell proliferation or for
controlling fertility. The medicaments can also be used to treat
hyperproliferative diseases. Fragments between amino acids 55 and 235
(preferably between amino acids 129 and 228) of a mammalian kin17 protein

CC are useful for regulating the interaction between proteins and curved
 CC DNA. The fragment can be used to block replication of HIV or its
 CC integration into the human genome or to target repair enzymes to curved
 CC DNA sites. Expression vectors for kin17 can be used for controlling cell
 CC proliferation

SQ Sequence 21 BP; 6 A; 5 C; 5 G; 5 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 2; Length 21;
 Best Local Similarity 100.0%; Pred. No. 1.7;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCAAGACAACTGTTGCTGGC 21
 |||||
 Db 1 TCAAGACAACTGTTGCTGGC 21

RESULT 2
 AAX85550
 ID AAX85550 standard; cDNA; 1102 BP.

XX AC AAX85550;
 XX DT 07-SEP-1999 (first entry)
 XX DE cDNA of a gene coding for a mouse deleted kin17 protein.

XX kin17 protein; cell proliferation; fertility; hyperproliferative disease;
 XX protein interaction; curved DNA; HIV replication; HIV integration;
 KW repair enzyme; ss.
 XX OS Mus sp.

XX FR2772046-A1.
 XX PN 11-JUN-1999.
 XX PF 09-DEC-1997; 97FR-00015536.
 XX PR 09-DEC-1997; 97FR-00015536.

XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
 XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
 XX WPI; 1999-359999/31.

XX New DNA coding for human kin17 protein - useful for controlling cell
 PT proliferation or fertility.
 XX Claim 4; Page 31; 69pp; French.

XX The present sequence encodes a mouse kin17 protein with amino acids 129-
 CC 228 deleted. The mammalian kin17 protein is useful for preparing a
 CC medicament for controlling cell proliferation or for controlling
 CC fertility. The medicaments can also be used to treat hyperproliferative
 CC diseases. Fragments between amino acids 55 and 235 (preferably between
 CC amino acids 129 and 228) of a mammalian kin17 protein are useful for
 CC regulating the interaction between proteins and curved DNA. The fragment
 CC can be used to block replication of HIV or its integration into the human
 CC genome or to target repair enzymes to curved DNA sites. Expression
 CC vectors for kin17 can be used for controlling cell proliferation
 XX
 XX SQ Sequence 1102 BP; 373 A; 205 C; 272 G; 252 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 2; Length 1102;
 Best Local Similarity 100.0%; Pred. No. 3.4;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCAAGACAACTGTTGCTGGC 21
 |||||
 Db 150 TCAAGACAACTGTTGCTGGC 170

RESULT 3
 AAX85570
 ID AAX85570 standard; cDNA; 1390 BP.

XX AC AAX85570;

XX DT 07-SEP-1999 (first entry)

XX DE cDNA of a gene coding for the murine kin17 protein.

XX Mouse; kin17 protein; cell proliferation; fertility;
 KW hyperproliferative disease; protein interaction; curved DNA;
 KW HIV replication; HIV integration; repair enzyme; ss.

XX OS Mus sp.

XX PN FR2772046-A1.

XX PD 11-JUN-1999.

XX PF 09-DEC-1997; 97FR-00015536.

XX PR 09-DEC-1997; 97FR-00015536.

XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
 XX WPI; 1999-359999/31.

XX New DNA coding for human kin17 protein - useful for controlling cell
 PT proliferation or fertility.

XX Claim 21; Page 35-36; 69pp; French.

XX The present sequence encodes a murine kin17 protein. The mammalian kin17
 CC protein is useful for preparing a medicament for controlling cell
 CC proliferation or for controlling fertility. The medicaments can also be
 CC used to treat hyperproliferative diseases. Fragments between amino acids
 CC 55 and 235 (preferably between amino acids 129 and 228) of a mammalian
 CC kin17 protein are useful for regulating the interaction between proteins
 CC and curved DNA. The fragment can be used to block replication of HIV or
 CC its integration into the human genome or to target repair enzymes to
 CC curved DNA sites. Expression vectors for kin17 can be used for
 CC controlling cell proliferation

XX SQ Sequence 1390 BP; 482 A; 256 C; 353 G; 299 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 2; Length 1390;
 Best Local Similarity 100.0%; Pred. No. 3.6;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCAAGACAACTGTTGCTGGC 21
 |||||
 Db 150 TCAAGACAACTGTTGCTGGC 170

RESULT 4
 AAQ79936
 ID AAQ79936 standard; cDNA; 1458 BP.

XX AC AAQ79936;

XX DT 25-MAR-2003 (revised)

XX DT 06-SEP-1995 (first entry)

XX DE Murine Kin17 cDNA.

XX chromosomal rearrangement; kin17 protein; SOS DNA repair system; RecA;
 KW genotoxic agent; zinc finger; DNA binding protein; ss.

XX OS Mus musculus.

XX Key Location/Qualifiers
FT primer_bind /*tag= b
FT /*note= "Oligo L (AAQ79946) binds to complementary strand"
FT 22. .1434
FT misc_feature /*tag= 1
FT /label= kin17 cDNA
FT /note= "nucleotides 1-1414; the genomic DNA contains at
FT least two introns within this sequence, see Comments"
FT 32. .49
FT primer_bind /*tag= c
FT /note= "Oligo C (AAQ79938) binding site"
FT 46. .1221
FT CDS /*tag= a
FT /product= "Kin17"
FT /note= "N's in the sequence denote illegible residues"
FT complement (67. .86)
FT primer_bind /*tag= d
FT /note= "Oligo S (AAQ79947) binds to complementary strand"
FT 274. .297
FT primer_bind /*tag= e
FT /note= "Oligo D (AAQ79939) binding site"
FT complement (339. .360)
FT primer_bind /*tag= f
FT /note= "Oligo K (AAQ79945) binds to complementary strand"
FT 451. .474
FT primer_bind /*tag= g
FT /note= "Oligo J (AAQ79944) binding site"
FT complement (550. .567)
FT primer_bind /*tag= h
FT /note= "Oligo E (AAQ79940) binds to complementary strand"
FT 802. .825
FT primer_bind /*tag= i
FT /note= "Oligo F (AAQ79941) binding site"
FT complement (839. .862)
FT primer_bind /*tag= j
FT /note= "Oligo G (AAQ79942) binds to complementary strand"
FT complement (1435. .1458)
FT primer_bind /*tag= k
FT /note= "Oligo B (AAQ79937) binds to complementary strand"
FR2706487-A1.
XX
XX
XX 23-DEC-1994.
XX
XX 15-JUN-1993; 93FR-00007171.
XX
XX 15-JUN-1993; 93FR-00007171.
XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
XX
XX Angulo-Mora JF, Tissier A, Frelat G, Mauffrey P, Guilly M;
XX WPI; 1995-039031/06.
XX
XX Purified murine kin17 protein prepn. for detecting chromosomal
XX rearrangements - also related antibodies, human and murine DNA, primers,
XX probes and vectors, used to assess damage caused by genotoxic agents.
XX
XX Claim 9; Page 33; 54pp; French.
XX
XX The murine Kin17 protein includes a zinc finger domain (see AAR66766),
XX recognises single- and double-stranded DNA (partic. regions of secondary
XX structure), has apparent mol. wt. 43 kD and is recognised by both anti-
XX kin17 antibodies and antibodies against the RecA protein of E.coli. The
XX Kin17 protein is involved in DNA repair; it can be used to monitor
XX chromosomal rearrangements following exposure to genotoxic agents. The
XX kin17 cDNA sequence AAQ79936 consists of a 1414 nucleotide sequence,
XX flanked by primer binding sites; the genomic kin17 DNA is claimed in
XX which an intron is inserted at position 137 of the 1414 nucleotide cDNA
XX and a second intron is located between nucleotides 339-429 of the cDNA.
XX (Updated on 25-MAR-2003 to correct PN field.)

XX
SQ Sequence 1458 BP; 496 A; 272 C; 372 G; 311 T; 0 U; 7 Other;
Query Match 100.0%; Score 21; DB 2; Length 1458;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCAAAGACAACTGTGTGCTGC 21
DB 195 TCAAAGACAACTGTGTGCTGC 215
RESULT 5
AAS27632/c
ID AAS27632 standard; DNA; 10319 BP.
XX
AC AAS27632;
XX
DT 07-NOV-2001 (first entry)
XX
DE DNA encoding novel signal transduction pathway protein, Seq ID 1292.
XX
KW Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;
KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;
KW immune system disorder; rheumatoid arthritis; inflammatory condition;
KW organ transplant rejection; infection; hepatitis C; blood disorder;
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
KW reproductive system; gastrointestinal; liver disorder; AIDS; ds;
KW acquired immune deficiency syndrome.
XX
OS Homo sapiens.
XX
FN WO200154733-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001312.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218220P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.


```
QY      2 CAAAGACAACTGTGCTGGC 21
Db      398 CAAAGACAACTGTGCTGGC 379

RESULT 6
ADB94435/C
ID      ADB94435 standard; DNA; 10319 BP.
XX
AC      ADB94435;
XX
DT      04-DEC-2003 (first entry)
XX
DE      Novel human protein:DNA #44.
XX
KW      ds; gene; human; autoimmune disease; Parkinson's disease; silicosis;
KW      gastrointestinal disease; atherosclerosis; haemophilia; thrombocytopenia;
KW      immunosuppressive agent; adjuvant; enhance immune response;
KW      higher affinity antibody induction;
KW      increased serum immunoglobulin concentration.
XX
OS      Homo sapiens.
XX
PN      US2002168711-A1.
XX
PD      14-NOV-2002.
XX
PF      17-JAN-2001; 2001US-00764868.
XX
PR      31-JAN-2000; 2000US-0179065P.
PR      04-FEB-2000; 2000US-0180628P.
PR      28-JUN-2000; 2000US-0214886P.
PR      07-JUL-2000; 2000US-0216647P.
PR      07-JUL-2000; 2000US-0216880P.
PR      11-JUL-2000; 2000US-0217487P.
PR      11-JUL-2000; 2000US-0217496P.
PR      14-JUL-2000; 2000US-0218290P.
PR      26-JUL-2000; 2000US-0220963P.
PR      26-JUL-2000; 2000US-0220964P.
PR      14-AUG-2000; 2000US-0224518P.
PR      14-AUG-2000; 2000US-0224519P.
PR      14-AUG-2000; 2000US-0225267P.
PR      14-AUG-2000; 2000US-0225268P.
PR      14-AUG-2000; 2000US-0225270P.
PR      14-AUG-2000; 2000US-0225447P.
PR      14-AUG-2000; 2000US-0225757P.
PR      14-AUG-2000; 2000US-0225758P.
PR      22-AUG-2000; 2000US-0226868P.
PR      30-AUG-2000; 2000US-0228924P.
PR      01-SEP-2000; 2000US-0229287P.
PR      01-SEP-2000; 2000US-0229343P.
PR      01-SEP-2000; 2000US-0229344P.
PR      01-SEP-2000; 2000US-0229345P.
PR      05-SEP-2000; 2000US-0229509P.
PR      05-SEP-2000; 2000US-0229513P.
PR      08-SEP-2000; 2000US-0231413P.
PR      21-SEP-2000; 2000US-0234223P.
PR      21-SEP-2000; 2000US-0234274P.
PR      25-SEP-2000; 2000US-0234997P.
PR      27-SEP-2000; 2000US-0235834P.
PR      29-SEP-2000; 2000US-0236327P.
PR      29-SEP-2000; 2000US-0236367P.
PR      29-SEP-2000; 2000US-0236368P.
PR      29-SEP-2000; 2000US-0236369P.
PR      29-SEP-2000; 2000US-0236370P.
PR      02-OCT-2000; 2000US-0236802P.
PR      02-OCT-2000; 2000US-0237037P.
PR      02-OCT-2000; 2000US-0237038P.
PR      02-OCT-2000; 2000US-0237039P.
PR      12-OCT-2000; 2000US-0237040P.
PR      13-OCT-2000; 2000US-0239935P.
PR      20-OCT-2000; 2000US-0240960P.
PR      20-OCT-2000; 2000US-0241785P.

20-OCT-2000; 2000US-0241809P.
01-NOV-2000; 2000US-0244617P.
17-NOV-2000; 2000US-0249299P.
08-DEC-2000; 2000US-0251856P.
08-DEC-2000; 2000US-0251868P.
08-DEC-2000; 2000US-0251869P.
(ROSE/) ROSEN C A.
(RUBE/) RUBEN S M.
(BARA/) BARASH S C.
Rosen CA, Ruben SM, Barash SC;
WPI; 2003-719985/68.
New isolated polypeptide useful for diagnosing and treating
immunosuppressive conditions such as autoimmune disease and Parkinson's
disease.
Disclosure; SEQ ID NO 1292; 345pp; English.
The invention relates to an isolated polypeptide. The polypeptide is
useful for diagnosing a pathological condition or a susceptibility to a
pathological condition in a subject, by determining the presence or
amount of expression of the polypeptide in a biological sample and
diagnosing a pathological condition or a susceptibility to a pathological
condition based on the presence or amount of expression of the
polypeptide. The polypeptide is also useful for identifying a binding
partner to the polypeptide, which involves contacting the polypeptide
with a binding partner and determining whether the binding partner
effects an activity of the polypeptide. The polypeptide or the nucleic
acid encoding the polypeptide is useful for preventing, treating, or
ameliorating a medical condition, which involves administering the
polypeptide or the nucleic acid to a mammalian subject. The nucleic acid
is useful for diagnosing a pathological condition or a susceptibility to
a pathological condition in a subject, which involves determining the
presence or absence of a mutation in the nucleic acid, and diagnosing a
pathological condition or susceptibility to a pathological condition
based on the presence or absence of the mutation. The polypeptide, the
nucleic acid and an antibody to the polypeptide are useful for treating
autoimmune disease, Parkinson's disease, silicosis, gastrointestinal
disease, atherosclerosis, haemophilia, thrombocytopenia. The polypeptide,
the nucleic acid and the antibody are useful as immunosuppressive agents,
as adjuvants to enhance immune responses, and as agents to induce higher
affinity antibodies and increase serum immunoglobulin concentrations. The
present sequence represents DNA encoding a novel human protein. Note: The
sequence data for this patent did not form part of the printed
specification but was obtained in electronic format direct from USPTO at
seqdata.uspto.gov/sequence.html?DocID=20020168711.
Sequence 10319 BP; 3036 A; 1953 C; 2147 G; 3183 T; 0 U; 0 Other;
Query Match      87.6%; Score 18.4; DB 10; Length 10319;
Best Local Similarity 95.0%; Pred. No. 99;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY      2 CAAAGACAACTGTGCTGGC 21
Db      398 CAAAGACAACTGTGCTGGC 379

RESULT 7
AAX85568
ID      AAX85568 standard; cDNA; 21 BP.
XX
AC      AAX85568;
XX
DT      07-SEP-1999 (first entry)
XX
DE      PCR primer and probe used to detect human kin17 nucleic acids.
XX
KW      Human; kin17 protein; cell proliferation; fertility; probe;
KW      hyperproliferative disease; protein interaction; curved DNA;
```


CC are useful for regulating the interaction between proteins and curved
CC DNA. The fragment can be used to block replication of HIV or its
CC integration into the human genome or to target repair enzymes to curved
CC DNA sites. Expression vectors for kin17 can be used for controlling cell
CC proliferation

SQ Sequence 1002 BP; 343 A; 189 C; 230 G; 240 T; 0 U; 0 Other;

Query Match 84.8%; Score 17.8; DB 2; Length 1002;
Best Local Similarity 90.5%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCAAAGACAACCTGTTGCTGGC 21
||| ||||| ||||| ||||| |||||
Db 207 TCAGAGACAACCTATTGCTGGC 227

RESULT 10
AAX85549
ID AAX85549 standard; cDNA; 1296 BP.

AC AAX85549;
DT 07-SEP-1999 (first entry)
XX cDNA of a gene coding for the human kin17 protein.

DE Human; kin17 protein; cell proliferation; fertility;
KW hyperproliferative disease; protein interaction; curved DNA;
KW HIV replication; HIV integration; repair enzyme; ss.

XX Homo sapiens.

OS FR2772046-A1.

PN 11-JUN-1999.

XX 09-DEC-1997; 97FR-00015536.

XX 09-DEC-1997; 97FR-00015536.

XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;

XX WPI; 1999-359999/31.

XX New DNA coding for human kin17 protein - useful for controlling cell

XX proliferation or fertility.

XX Claim 1; Page 30-31; 69pp; French.

XX The present sequence encodes a human kin17 protein. The mammalian kin17
CC protein is useful for preparing a medicament for controlling cell
CC proliferation or for controlling fertility. The medicaments can also be
CC used to treat hyperproliferative diseases. Fragments between amino acids
CC 55 and 235 (preferably between amino acids 129 and 228) of a mammalian
CC kin17 protein are useful for regulating the interaction between proteins
CC and curved DNA. The fragment can be used to block replication of HIV or
CC its integration into the human genome or to target repair enzymes to
CC curved DNA sites. Expression vectors for kin17 can be used for
CC controlling cell proliferation

SQ Sequence 1296 BP; 453 A; 240 C; 310 G; 293 T; 0 U; 0 Other;

Query Match 84.8%; Score 17.8; DB 2; Length 1296;
Best Local Similarity 90.5%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCAAAGACAACCTGTTGCTGGC 21
||| ||||| ||||| ||||| |||||
Db 207 TCAGAGACAACCTATTGCTGGC 227

RESULT 11
ABK42300
ID ABK42300 standard; DNA; 12555 BP.

XX AC ABK42300;

XX 21-MAY-2002 (first entry)

XX Genomic sequence #199 encoding novel human connective tissue polypeptide.
DE Human; connective tissue related disorder; cancer; gene therapy;
KW cytoskeletal; gene; ds.

XX Homo sapiens.

XX WO200155343-A1.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US001322.

XX 31-JAN-2000; 2000US-0179065P.

XX 04-FEB-2000; 2000US-0180628P.

XX 24-FEB-2000; 2000US-0184664P.

XX 02-MAR-2000; 2000US-0186350P.

XX 16-MAR-2000; 2000US-0189874P.

XX 17-MAR-2000; 2000US-0190076P.

XX 18-APR-2000; 2000US-0198123P.

XX 19-MAY-2000; 2000US-0205515P.

XX 07-JUN-2000; 2000US-0209467P.

XX 28-JUN-2000; 2000US-0214886P.

XX 30-JUN-2000; 2000US-0215135P.

XX 07-JUL-2000; 2000US-0216647P.

XX 07-JUL-2000; 2000US-0216800P.

XX 11-JUL-2000; 2000US-0217487P.

XX 14-JUL-2000; 2000US-0218290P.

XX 26-JUL-2000; 2000US-0220963P.

XX 26-JUL-2000; 2000US-0220964P.

XX 14-AUG-2000; 2000US-0224518P.

XX 14-AUG-2000; 2000US-0224519P.

XX 14-AUG-2000; 2000US-0225213P.

XX 14-AUG-2000; 2000US-0225214P.

XX 14-AUG-2000; 2000US-0225266P.

XX 14-AUG-2000; 2000US-0225267P.

XX 14-AUG-2000; 2000US-0225268P.

XX 14-AUG-2000; 2000US-0225447P.

XX 14-AUG-2000; 2000US-0225757P.

XX 14-AUG-2000; 2000US-0225758P.

XX 18-AUG-2000; 2000US-0226279P.

XX 22-AUG-2000; 2000US-0226681P.

XX 22-AUG-2000; 2000US-0226686P.

XX 22-AUG-2000; 2000US-0227182P.

XX 23-AUG-2000; 2000US-0227009P.

XX 30-AUG-2000; 2000US-0228924P.

XX 01-SEP-2000; 2000US-0229287P.

XX 01-SEP-2000; 2000US-0229343P.

XX 01-SEP-2000; 2000US-0229344P.

XX 01-SEP-2000; 2000US-0229345P.

XX 05-SEP-2000; 2000US-0229509P.

XX 05-SEP-2000; 2000US-0229513P.

XX 06-SEP-2000; 2000US-0230437P.

XX 06-SEP-2000; 2000US-0230438P.

XX 08-SEP-2000; 2000US-0231242P.

XX 08-SEP-2000; 2000US-0231243P.

XX 08-SEP-2000; 2000US-0231244P.

XX 08-SEP-2000; 2000US-0231413P.

XX 08-SEP-2000; 2000US-0231414P.

XX 08-SEP-2000; 2000US-0232080P.

XX 08-SEP-2000; 2000US-0232081P.

```
PR 12-SEP-2000; 2000US-02311968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-565190/63.
XX
XX Nucleic acid encoding novel connective tissue associated polypeptides,
XX used in diagnosing, preventing, treating or ameliorating a disorder such
XX as cancer or rheumatoid arthritis.
XX
XX Disclosure; SEQ ID NO 1187; 673pp; English.
XX
XX The present invention relates to the isolation of novel human connective
XX tissue related polypeptides (AAU86435-AAU86923) and the polynucleotide
XX (cDNA and genomic) sequences encoding them. The sequences of the
XX invention are useful in the diagnosis, treatment, prevention and/or
XX prognosis of diseases associated with connective tissue(s), including
XX cancer. The polynucleotide sequences of the invention are also useful in
XX gene therapy. ABK42102-ABK43116 represent genomic sequences encoding the
XX novel human connective tissue related polypeptides. Note: The sequence
XX data for this patent did not form part of the printed specification, but
XX was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 12555 BP; 2632 A; 3328 C; 3392 G; 3203 T; 0 U; 0 Other;
XX
XX Query Match 84.8%; Score 17.8; DB 4; Length 12555;
XX Best Local Similarity 90.5%; Pred. No. 2e+02;
XX Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
QY 1 TCAAAGACAACTGTGTGCTGGC 21
Db 8515 TCAAAGACAGCTTTTGTCTGGC 8535
XX
RESULT 12
AAK78331
ID AAK78331 standard; DNA; 12555 BP.
XX
AC AAK78331;
XX
XX 07-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:33143.
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001354.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
```


CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 12555 BP; 2632 A; 3328 C; 3392 G; 3203 T; 0 U; 0 Other;

Query Match 84.8%; Score 17.8; DB 4; Length 12555;
Best Local Similarity 90.5%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCAGAGACAACTGTTGCTGCGC 21

Db 8515 TCAGAGACAGCTTTGCTGCGC 8535

RESULT 13

ID ADB60456

XX ADB60456 standard; DNA; 12555 BP.

AC ADB60456;

XX 04-DEC-2003 (first entry)

DE Connective tissue related genomic DNA #199.

XX cytotatic; neuroprotective; nootropic; antiparkinsonian; cardiovascular;

KW antiarteriosclerotic; immunosuppressive; antirheumatic; antiarthritic;
KW antiinflammatory; antiallergic; antiasthmatic; dermatological;
KW nephrotropic; virucide; fungicide; antibacterial; antiparasitic;
KW gene therapy; ds; connective tissues disorder; rheumatoid arthritis;
KW systemic lupus erythematosus; scleroderma; Sjogren's syndrome; cancer;
KW cancer metastasis; neoplasia; leukaemia; neurodegenerative disorder;
KW Alzheimer's disease; Parkinson's disease; cardiovascular disease;
KW atherosclerosis; myocarditis; cardiopulmonary bypass complication;
KW autoimmune disease; multiple sclerosis; allergic reaction; asthma;
KW rhinitis; eczema; inflammatory condition; Crohn's disease; nephritis;
KW gastrointestinal disorder; inflammatory bowel disease;
KW organ transplant rejection; immune system disorder; Bruton's disease;
KW X-linked lymphoproliferative syndrome;
KW B-cell lymphoproliferative disorder; HIV; AIDS; infection;
KW chromosome identification; chromosome mapping;
KW connective tissue related polynucleotide; gene; ds.

OS Homo sapiens.

XX US2003054375-A1.

PN 20-MAR-2003.

PD 07-MAR-2002; 2002US-00092154.

XX 31-JAN-2000; 2000US-0179065P.

XX 04-FEB-2000; 2000US-0180628P.

XX 24-FEB-2000; 2000US-0184564P.

XX 02-MAR-2000; 2000US-0186350P.

XX 16-MAR-2000; 2000US-0189874P.

XX 17-MAR-2000; 2000US-0190076P.

XX 18-APR-2000; 2000US-0198123P.

XX 19-MAY-2000; 2000US-0205515P.

XX 07-JUN-2000; 2000US-0209467P.

XX 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0232081P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.

PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254037P.
PR 05-JAN-2001; 2001US-0259678P.
PR 17-JAN-2001; 2001US-00764847.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Ruben SM, Barash SC;
XX
XX WPI; 2003-634869/60.
XX P-PSDB; ADB59396.
XX
XX New connective tissue-related polypeptides and polynucleotides, useful
PT for treating, preventing and/or prognosing e.g. disorders of connective
PT tissue, (e.g. rheumatoid arthritis), cancers, cancer metastases and/or
PT neoplasias.
XX
XX Disclosure; SEQ ID NO 1187; 248pp; English.
PS
XX
XX The invention describes an isolated nucleic acid molecule (I), which
CC comprises a sequence that is at least 95 % identical to a connective
CC tissue-related polynucleotide encoding connective tissue antigens (CTA).
CC The polypeptide or polynucleotide is useful for preventing, treating, or

CC ameliorating medical conditions in a mammal. The connective tissue
CC polypeptides, polynucleotides and antibodies are particularly useful for
CC treating, preventing and/or prognosing disorders of connective tissues
CC (e.g. rheumatoid arthritis, discoid and systemic lupus erythematosus,
CC scleroderma, or Sjogren's syndrome), cancers, cancer metastases and/or
CC neoplasias (e.g. leukaemia), neurodegenerative disorders (e.g.
CC Alzheimer's disease, or Parkinson's disease), cardiovascular diseases
CC (e.g. atherosclerosis, myocarditis or cardiopulmonary bypass
CC complications), autoimmune diseases (e.g. systemic lupus erythematosus,
CC rheumatoid arthritis, or multiple sclerosis), allergic reactions (e.g.
Query Match 84.8%; Score 17.8; DB 9; Length 12555;
Best Local Similarity 90.5%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TCAAGACAACTGTTGCTGSC 21
DB 8515 TCAAGACAGACTTTTCTGCGC 8535
RESULT 14
ABL06020
ID ABL06020 standard; cDNA; 13810 BP.
XX
AC ABL06020;
XX
XX 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 12542.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX
XX WPI; 2001-656860/75.
XX P-PSDB; ABB61917.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
XX
XX Claim 1; SEQ ID NO 12542; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX
XX Sequence 13810 BP; 4122 A; 2676 C; 2815 G; 4197 T; 0 U; 0 Other;
Query Match 84.8%; Score 17.8; DB 4; Length 13810;
Best Local Similarity 90.5%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY

1

TCAAAGACAACTGTTGCTGGC 21

Db

9865

TTAAAGCCAACTGTTGCTGGC 9865

RESULT 15

ACF85421

ID ACF85421 standard; DNA; 329 BP.

XX

AC

ACF85421;

XX

DT

02-JUN-2005 (first entry)

XX

DE

Human SIRS/sepsis diagnostic marker DNA fragment 4281.

XX

KW

Systemic inflammatory response syndrome; SIRS; antibacterial;

KW

immunosuppressive; antiinflammatory; diagnosis; sepsis; ds.

XX

OS

Homo sapiens.

XX

FN

WO2004087949-A2.

XX

PD

14-OCT-2004.

XX

PF

31-MAR-2004; 2004WO-EP003419.

XX

PR

02-APR-2003; 2003DE-01015031.

PR

08-AUG-2003; 2003DE-01036511.

PR

02-SEP-2003; 2003DE-01040395.

XX

PA

(SIRS-) SIRS LAB GMBH.

XX

PI

Rusewurm S, Reinhart K, Saluz H, Straube E, Zipfel PF, Deigner H;

XX

DR

WPI; 2004-748070/73.

XX

PT

In vitro detection of systemic inflammatory response syndrome and related

PT

conditions, for e.g. monitoring progression, comprises detecting abnormal

PT

expression of disease-related genes.

XX

PS

Disclosure; Page; 75pp; German.

XX

CC

The invention relates to a novel method for in vitro detection of

CC

systemic inflammatory response syndrome (SIRS). The method comprises

CC

detecting abnormal expression of disease-related genes, or their

CC

associated peptides. The method of the invention demonstrates

CC

antibacterial, immunosuppressive and antiinflammatory applications and

CC

may be used for early differential diagnosis, monitoring progression,

CC

assessing risk, assessing the likely response to treatment and for post

CC

mortem diagnosis of systemic inflammatory response syndrome, sepsis and

CC

sepsis-like conditions. The recombinant or synthetic nucleic acid

CC

sequences of the invention, or derived proteins or peptides, may be

CC

useful as calibrants in assays for the specified diseases, for evaluating

CC

activity or toxicity in screening for active agents and/or for

CC

preparation of agents for treatment or prevention of the specified

CC

diseases. The current sequence is that of a human SIRS/sepsis diagnostic

CC

marker DNA fragment of the invention. Note: The sequence data for this

CC

patent did not form part of the printed specification, but was obtained

CC

in electronic format directly from WIPO at ftp.wipo.int/pub/published

CC

pct_sequences. Furthermore, a number of arbitrary SEQ ID NO.8 are

CC

disclosed within the specification, however, these have not been taken

CC

into account during indexing due to inconsistencies in application and

CC

format

XX

Sequence 329 BP; 111 A; 54 C; 73 G; 91 T; 0 U; 0 Other;

Query Match

80.0%;

Score 16.8;

DB 13;

Length 329;

Best Local Similarity

90.0%;

Pred. No. 3.5e+02;

Matches 18;

Conservative 0;

Mismatches 2;

Indels 0;

Gaps 0;

QY

2

CAAAGACAACTGTTGCTGGC 21

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November '26, 2005, 12:58:53 ; Search time 117.387 Seconds
(without alignments)
8370.011 Million cell updates/sec

Title: US-09-555-529-27

Perfect score: 21
Sequence: 1 tcaaaqcaactgttctgctggc 21

Scoring table: IDENTITY_NUC
Gapop 10'0 , Gapext 1.0

Searched: 41078325 [seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*

1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_hic.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_est7.*
9: gb_gss1.*
10: gb_gss2.*
11: gb_gss3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	355	5	BY038159
2	21	100.0	407	1	AA684460 vm64d01.8
3	21	100.0	465	1	A1180541 uc70c09.r
4	21	100.0	496	2	B8665972 BB866972
5	21	100.0	500	7	CF916748 CF916748
6	21	100.0	501	6	CA560899 K0278H05-
7	21	100.0	515	1	AA638405 v199f02.r
8	21	100.0	538	6	CA559106 K0254C06-
9	21	100.0	573	7	CF914584 CF914584
10	21	100.0	583	2	BF321725 BF321725
11	21	100.0	600	5	B0919249 B0919249
12	21	100.0	625	7	CF914851 CF914851
13	21	100.0	638	4	AK009429 AK009429
14	21	100.0	653	6	CF172363 CF172363
15	21	100.0	653	6	CF172995 CF172995
16	21	100.0	654	2	BE226431 BE226431
17	21	100.0	660	6	CB847652 CB847652
18	21	100.0	663	8	CX227787 CX227787
19	21	100.0	674	7	CF914272 CF914272
20	21	100.0	696	6	CF174012 CF174012
21	21	100.0	740	8	CX763426 CX763426
22	21	100.0	942	5	BQ714508 BQ714508

23	21	100.0	1077	5	BUS11663
24	19.4	92.4	342	8	DN131260
25	19.4	92.4	377	8	DN133257
26	19.4	92.4	440	6	CB749078
27	19.4	92.4	454	8	DN127913
28	19.4	92.4	487	2	BF396990
29	19.4	92.4	512	2	BE014334
30	19.4	92.4	544	6	CB288055
31	19.4	92.4	630	5	BX924377
32	19.4	92.4	635	5	BX925174
33	19.4	92.4	645	7	CN166982
34	19.4	92.4	645	8	DN128273
35	19.4	92.4	645	8	DN133982
36	19.4	92.4	676	2	BI110637
37	19.4	92.4	723	7	CK474145
38	19.4	92.4	783	7	CN166221
39	19.4	92.4	1543	4	AY609991
40	19.4	90.5	531	7	CO061759
41	18.4	87.6	406	9	CE014061
42	18.4	87.6	419	10	CL809118
43	18.4	87.6	488	10	CL736559
44	18.4	87.6	535	10	CL726793
45	18.4	87.6	612	9	AQ288077

ALIGNMENTS

RESULT 1
BY038159
LOCUS
DEFINITION
ACCSSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

BY038159 RIKEN full-length enriched, cDNA clone 1730004D10 5', mRNA sequence.
355 bp mRNA linear EST 06-DEC-2002
BY038159.1 GI:26143602
EST.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
1 (bases 1 to 355)
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaide, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Haegawa, Y., Nogami, A., Schonbach, C., Gojibori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Bruscia, V., Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Glasi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawai, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Kongaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

TITLE
JOURNAL
PUBMED

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
12466851

JOURNAL COMMENT	Unpublished (1996) Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of MedicineP 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.		
	MG1:567225 Possible reversed clone: similarity on wrong strand High quality sequence stop: 370.		
FEATURES	source	Location/Qualifiers 1. .407 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J x DBA/2J F1" /db_xref="taxon:10090" /clone="IMAGE:1003009" /tissue_type="embryo" /dev_stage="2-cell" /lab_host="DH10B" /clone_lib="Knowles Solter mouse 2 cell" /note="Organ: embryo; Vector: pBluescribe (modified); Site 1: MluI; Site 2: SalI; Cloned unidirectionally from mRNA prepared from 13,500 2-cell stage embryos. Primer: Sali (dtr): 5'-CGGTGACCGTCGACCGTTTTTTTTTT-3'. cDNAs were cloned into the MluI/Sali sites of a modified pBluescribe vector using commercial linkers (NEB). Average insert size: 1.2 kb."	
		ORIGIN Query Match 100.0%; Score 21; DB 1; Length 407; Best Local Similarity 100.0%; Pred. No. 18; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0; QY 1 TCAAAGACAACTGTTGCTGGC 21 Db 91 TCAAAGACAACTGTTGCTGGC 111	
RESULT 3	AL180541 LOCUS DEFINITION	465 bp mRNA linear EST 08-OCT-1998 uc70c09.r1 Soares mammary_gland_NbMWG Mus musculus cDNA clone IMAGE:1430992 5' similar to SW:RTS2_YEAST P40962 ZINC FINGER PROTEIN RTS2. ;, mRNA sequence. AI180541 AI180541.1 GI:3731179 Mus musculus (house mouse)	
		ACCESSION VERSION KEYWORDS SOURCE ORGANISM Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus. 1 (bases 1 to 465) Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R. The WashU-HHMI Mouse EST Project Unpublished (1996) Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of MedicineP 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.	
REFERENCE	AUTHORS	TITLE JOURNAL COMMENT	
		Unpublished (1996) Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of MedicineP 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.	

JOURNAL COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gs.riken.jp, URL:http://genome.gsc.riken.jp/ Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata M., Nakamura M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001) Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site (http://genome.gsc.riken.go.jp) for further details.		
	Location/Qualifiers 1. .355 /organism="Mus musculus" /mol_type="mRNA" /strain="DBA/2" /db_xref="taxon:10090" /clone="I730004D10" /cell_line="CRL-1722 L5178Y-R" /clone_lib="RIKEN full-length enriched, CRL-1722 L5178Y-R"		
FEATURES	source	1. .355 /organism="Mus musculus" /mol_type="mRNA" /strain="DBA/2" /db_xref="taxon:10090" /clone="I730004D10" /cell_line="CRL-1722 L5178Y-R" /clone_lib="RIKEN full-length enriched, CRL-1722 L5178Y-R"	
		ORIGIN Query Match 100.0%; Score 21; DB 5; Length 355; Best Local Similarity 100.0%; Pred. No. 18; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0; QY 1 TCAAAGACAACTGTTGCTGGC 21 Db 173 TCAAAGACAACTGTTGCTGGC 193	
RESULT 2	AA684460 LOCUS DEFINITION	407 bp mRNA linear EST 09-DEC-1997 vm64d01.s1 Knowles Solter mouse 2 cell Mus musculus cDNA clone IMAGE:1003009 5' similar to SW:RTS2_YEAST P40962 RTS2 PROTEIN. [1] ;, mRNA sequence. AA684460 GI:2671046 Mus musculus (house mouse)	
		ACCESSION VERSION KEYWORDS SOURCE ORGANISM Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus. 1 (bases 1 to 407) Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R. The WashU-HHMI Mouse EST Project	
REFERENCE	AUTHORS	TITLE JOURNAL COMMENT	
		Unpublished (1996) Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.	

JOURNAL COMMENT

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:567225

Possible reversed clone: similarity on wrong strand

High quality sequence stop: 370.

Location/Qualifiers

source

1. .407

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J x DBA/2J F1"

/db_xref="taxon:10090"

/clone="IMAGE:1003009"

/tissue_type="embryo"

/dev_stage="2-cell"

/lab_host="DH10B"

/clone_lib="Knowles Solter mouse 2 cell"

/note="Organ: embryo; Vector: pBluescribe (modified);

Site: 1: MluI; Site 2: SalI; Cloned unidirectionally from

mRNA prepared from 13,500 2-cell stage embryos. Primer:

SalI (dT): 5'-CGGTGACGTCGACCGTTTTTTTTTT-3'. cDNAs

were cloned into the MluI/SalI sites of a modified

pBluescribe vector using commercial linkers (NEB).

Average insert size: 1.2 kb."

ORIGIN

Query Match 100.0%; Score 21; DB 1; Length 407;

Best Local Similarity 100.0%; Pred. No. 18;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGGC 21

|||||

Db 91 TCAAAGACAACTGTTGCTGGC 111

RESULT 3

AL180541

LOCUS

DEFINITION

AI180541 465 bp mRNA linear EST 08-OCT-1998
uc70c09.r1 Soares mammary_gland_NDMMG Mus musculus cDNA clone
IMAGE:1430992 5' similar to SW:RTS2_YEAST P40962 ZINC FINGER
PROTEIN RTS2. ;, mRNA sequence.

AI180541 GI:3731179

EST.

Mus musculus (house mouse)

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
1 (bases 1 to 465)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

The WashU-HHMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

TITLE

JOURNAL

COMMENT

CF916748	500 bp	mRNA	linear	EST 05-NOV-2003
LOCUS				
DEFINITION	B0999H08-5	NIA Mouse Unfractilized Egg cDNA Library (Long 1) Mus musculus cDNA clone NIA:B0999H08 IMAGE:30481819 5', mRNA sequence.		
ACCESSION	CF916748			

```

VERSION      CF916748.1  GI:38187950
KEYWORDS     EST.
SOURCE       Mus musculus (house mouse)
ORGANISM     Mus musculus
REFERENCE     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS       Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
TITLE         Sciurognathi; Muridae; Muridae; Murinae; Mus.
JOURNAL       1 (bases 1 to 500)
PUBMED       Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.
COMMENT       Construction of long-transcript enriched cDNA libraries from
              submicrogram amounts of total RNAs by a universal PCR amplification
              method
              Genome Res. 11 (9), 1553-1558 (2001)
              11541199
              Contact: Dawood B. Dudekula
              Laboratory of Genetics
              National Institute on Aging/National Institutes of Health
              333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
              Email: cdna@lgsun.grc.nia.nih.gov
              Plate: B0999 row: H column: 08
              Seq primer: M13 Reverse
              High quality sequence stop: 500
              POLYA=No.

FEATURES             Location/Qualifiers
     source           1..500
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="niaEST:B0999H08-5"
                     /db_xref="taxon:10090"
                     /clone="NIA:B0999H08 IMAGE:30481819"
                     /dev_stage="Unfertilized Egg"
                     /lab_host="DH10B"
                     /clone_lib="NIA Mouse Unfertilized Egg cDNA Library (Long
                     1)"
                     /note="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI;
                     Site 2: NotI; Mouse cDNA project by the Laboratory of
                     Genetics, National Institute on Aging (NIA), Intramural
                     Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA).
                     This is a long-transcript enriched cDNA library (Ref.
                     Genome Res. 11: 1553-1558 (2001). [PMID: 11541199]). Total
                     RNAs were extracted from a pool of 1488 unfertilized eggs.
                     Double-stranded cDNAs were synthesized with an Oligo(dT)
                     primer [Invitrogen]:
                     5'-pGACTAGTCTAGATCGAGCGCGCCGCTTTT-3',
                     treated with T4 DNA polymerase, and purified by
                     ethanol-precipitation. The cDNAs were ligated to
                     Lone-linker LL-Sal4, purified by phenol/chloroform, and
                     separated from free linkers by Centricon 100. Then, the
                     cDNAs were amplified by long-range high fidelity PCR using
                     Ex Taq polymerase (Takara) with a primer Sal4-S. The
                     products were purified by phenol/chloroform and Centricon
                     100. The cDNAs were digested with SalI and NotI enzymes
                     and cloned into SalI/NotI site of pCMV-SPORT6 plasmid
                     vector. The DH10B E. coli host was transformed with the
                     ligation mixture by the standard chemical method. The
                     average insert size is about 2.5 kb. The library was
                     constructed by Yulan Piao."

ORIGIN
Query Match      100.0%; Score 21; DB 7; Length 500;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAAGACAACTGTTGCTGGC 21
    |||
Db 136 TCAAGACAACTGTTGCTGGC 156

RESULT 6
CA560899
LOCUS      K0278H05-5N NIA Mouse Unfertilized Egg cDNA Library (Long) Mus
DEFINITION

```

```

musculus cDNA clone NIA:K0278H05 IMAGE:30052216 5', mRNA sequence.
CA560899
EST.
CA560899.1  GI:25105554
SOURCE       Mus musculus (house mouse)
ORGANISM     Mus musculus
REFERENCE     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS       Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
TITLE         Sciurognathi; Muridae; Muridae; Murinae; Mus.
JOURNAL       1 (bases 1 to 501)
PUBMED       Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Luo, A., Stagg, C.A.,
              Martin, P., Alba, K., Tanaka, T. and Ko, M.S.H.
              Systematic Analyses of NIA Mouse Unfertilized Egg cDNA Library
              (Long)
              Unpublished (2001)
              Other ESTs: K0278H05-3
              Contact: Dawood B. Dudekula
              Laboratory of Genetics
              National Institute on Aging/National Institutes of Health
              333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
              Email: cdna@lgsun.grc.nia.nih.gov
              Plate: K0278 row: H column: 05
              Seq primer: M13 Reverse
              High quality sequence stop: 501
              POLYA=No.

FEATURES             Location/Qualifiers
     source           1..501
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="niaEST:K0278H05-5N"
                     /db_xref="taxon:10090"
                     /clone="NIA:K0278H05 IMAGE:30052216"
                     /tissue_type="Unfertilized Egg"
                     /lab_host="DH10B"
                     /clone_lib="NIA Mouse Unfertilized Egg cDNA Library
                     (Long)"
                     /note="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2:
                     NotI; Mouse cDNA project by the Laboratory of Genetics,
                     National Institute on Aging (NIA), Intramural Research
                     Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is
                     a long-transcript enriched cDNA library (Ref. Genome Res.
                     11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
                     extracted from a pool of 1488 unfertilized eggs
                     Double-stranded cDNAs were synthesized with an Oligo(dT)
                     primer [Invitrogen]:
                     5'-pGACTAGTCTAGATCGAGCGCGCCGCTTTT-3',
                     treated with T4 DNA polymerase, and purified by
                     ethanol-precipitation. The cDNAs were ligated to
                     Lone-linker LL-Sal4, purified by phenol/chloroform, and
                     separated from free linkers by Centricon 100. Then, the
                     cDNAs were amplified by long-range high fidelity PCR using
                     Ex Taq polymerase (Takara) with a primer Sal4-S. The
                     products were purified by phenol/chloroform and Centricon
                     100. The cDNAs were digested with SalI and NotI enzymes
                     and cloned into SalI/NotI site of pSPORT1 plasmid vector.
                     The DH10B E. coli host was transformed with the ligation
                     mixture by the standard chemical method. The average
                     insert size is about 2.5 kb. The library was constructed
                     by Yulan Piao (NIA)."

ORIGIN
Query Match      100.0%; Score 21; DB 6; Length 501;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAAGACAACTGTTGCTGGC 21
    |||
Db 146 TCAAGACAACTGTTGCTGGC 166

RESULT 7
AA638405

```



```

LOCUS      AA638405      515 bp      mRNA      linear      EST 23-OCT-1997
DEFINITION v199f02.r1 Knowles Solter mouse blastocyst B1 Mus musculus cDNA
            clone IMAGE:98827 5' similar to SW:RTS2_YEAST_F40962 RTS2 PROTEIN.
            ; mRNA sequence.
ACCESSION  AA638405
VERSION    AA638405.1 GI:2561993
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 515)
AUTHORS    Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Stepcoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
TITLE      The WashU-HMI Mouse EST Project
JOURNAL    Unpublished (1996)
COMMENT    Contact: Marra M/Mouse EST Project
            WashU-HMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMGE Consortium (info@image.llnl.gov) for further information.
            MGI:561107
            High quality sequence stop: 457.
FEATURES   Location/Qualifiers
            source          1..515
                        /organism="Mus musculus"
                        /mol_type="mRNA"
                        /strain="C57BL/6J x DBA/2J F1"
                        /db_xref="taxon:10090"
                        /clone="IMAGE:98827"
                        /tissue_type="blastocyst"
                        /dev_stage="embryo (pre-implantation)"
                        /lab_host="DH10B"
                        /clone_lib="Knowles Solter mouse blastocyst B1"
                        /notes="Organ: embryo; Vector: pSPORT; Site 1: NotI;
                        Site 2: SalI; Cloned unidirectionally from mRNA prepared
                        from 800 blastocysts. Primer: SalI(dT):
                        5'-CGTCGACCGTCGACCGTTTTTTTTTT-3'. cDNAs were
                        cloned into the NotI/SalI sites of a pSPORT vector (Life
                        Technologies). Two different size selections: B1 (larger
                        inserts) and B3."
ORIGIN
            Query Match      100.0%; Score 21; DB 1; Length 515;
            Best Local Similarity 100.0%; Pred. No. 19;
            Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TCAAAGACAACACTGTGCTGGC 21
    |||||
Db 147 TCAAAGACAACACTGTGCTGGC 167

RESULT 8
CA559106      538 bp      mRNA      linear      EST 19-NOV-2002
LOCUS        K0254C06-5N NIA Mouse Unfertilized Egg cDNA Library (Long) Mus
DEFINITION   musculus cDNA clone NIA:K0254C06 IMAGE:30049853 5', mRNA sequence.
ACCESSION    CA559106
VERSION      CA559106.1 GI:25103669
KEYWORDS     EST.
SOURCE       Mus musculus (house mouse)
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.

LOCUS      AA638405      515 bp      mRNA      linear      EST 23-OCT-1997
DEFINITION v199f02.r1 Knowles Solter mouse blastocyst B1 Mus musculus cDNA
            clone IMAGE:98827 5' similar to SW:RTS2_YEAST_F40962 RTS2 PROTEIN.
            ; mRNA sequence.
ACCESSION  AA638405
VERSION    AA638405.1 GI:2561993
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE  1 (bases 1 to 538)
AUTHORS    Piao,Y., Kargul,G.J., Dudekula,D.B., Qian,Y., Luo,A., Stagg,C.A.,
            Martin,P., Aiba,K., Tanaka,T. and Ko,M.S.H.
TITLE      Systematic Analyses of NIA Mouse Unfertilized Egg cDNA Library
            (Long)
JOURNAL    Unpublished (2001)
COMMENT    Other ESTs: K0254C06-3
            Contact: Dawood B. Dudekula
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgun.grc.nia.nih.gov
            Plate: K0254 row: C column: 06
            Seq primer: M13 Reverse
            High quality sequence stop: 538
            POLYA=No.
FEATURES   Location/Qualifiers
            source          1..538
                        /organism="Mus musculus"
                        /mol_type="mRNA"
                        /strain="C57BL/6J"
                        /db_xref="niaEST:K0254C06-5N"
                        /db_xref="taxon:10090"
                        /clone="NIA:K0254C06 IMAGE:30049853"
                        /tissue_type="Unfertilized Egg"
                        /lab_host="DH10B"
                        /clone_lib="NIA Mouse Unfertilized Egg cDNA Library
                        (Long)"
                        /notes="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2:
                        NotI; Mouse cDNA project by the Laboratory of Genetics,
                        National Institute on Aging (NIA), Intramural Research
                        Program, NIH (http://lgun.grc.nia.nih.gov/cDNA). This is
                        a long-transcript enriched cDNA library (Ref. Genome Res.
                        11: 1553-1558 (2001)). [PMID: 11544199]. Total RNAs were
                        extracted from a pool of 1488 unfertilized eggs.
                        Double-stranded cDNAs were synthesized with an Oligo(dT)
                        primer [Invitrogen]:
                        5'-pGACTAGTCTAGTCGCGAGCGCGCCCTTTTTTTTTTT-3',
                        treated with T4 DNA polymerase, and purified by
                        ethanol-precipitation. The cDNAs were ligated to
                        Lone-linker LL-Sal4, purified by phenol/chloroform, and
                        separated from free linkers by Centricon 100. Then, the
                        cDNAs were amplified by long-range high fidelity PCR using
                        Ex Tag polymerase (Takara) with a primer Sal4-S. The
                        products were purified by phenol/chloroform and Centricon
                        100. The cDNAs were digested with SalI and NotI enzymes
                        and cloned into SalI/NotI site of pSPORT1 plasmid vector.
                        The DH10B E. coli host was transformed with the ligation
                        mixture by the standard chemical method. The average
                        insert size is about 2.5 kb. The library was constructed
                        by Yulan Piao (NIA)."
ORIGIN
            Query Match      100.0%; Score 21; DB 6; Length 538;
            Best Local Similarity 100.0%; Pred. No. 19;
            Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TCAAAGACAACACTGTGCTGGC 21
    |||||
Db 146 TCAAAGACAACACTGTGCTGGC 166

RESULT 9
CF914584      573 bp      mRNA      linear      EST 05-NOV-2003
LOCUS        B0966B12-5 NIA Mouse Unfertilized Egg cDNA Library (Long) Mus
DEFINITION   musculus cDNA clone NIA:B0966B12 IMAGE:30478583 5', mRNA sequence.
ACCESSION    CF914584
VERSION      CF914584.1 GI:38195786
KEYWORDS     EST.
SOURCE       Mus musculus (house mouse)
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```


Tel: 713 792 3646
Fax: 713 790 0329

FEATURES

source
1. .600
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/tissue_type="neural retina"
/dev_stage="embryonic day 14.5 post-fertilization"
/clone_lib="Mouse E14.5 retina lambda ZAP II Library"

ORIGIN

Query Match 100.0%; Score 21; DB 5; Length 600;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGCGC 21
|||||
Db 188 TCAAAGACAACTGTTGCTGCGC 208

RESULT 12

CF914851
LOCUS B0970B07-5 NIA Mouse Unfertilized Egg cDNA Library (Long 1) Mus
DEFINITION musculus cDNA clone NIA:B0970B07 IMAGE:30478962 5', mRNA sequence.

ACCESSION CF914851
VERSION CF914851.1 GI:38186053
KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidae; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 625)
Piao Y., Ko N.T., Lim M.K. and Ko M.S.H.
AUTHORS Construction of long-transcript enriched cDNA libraries from
TITLE submicrogram amounts of total RNAs by a universal PCR amplification
method

JOURNAL Genome Res. 11 (9), 1553-1558 (2001)

PUBMED 11544199

COMMENT Contact: Dawood B. Dudekula

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdna@igsun.grc.nia.nih.gov

Plate: B0970 row: B column: 07

Seq primer: M13 Reverse

High quality sequence stop: 625

POLYA=No.

FEATURES

source

Location/Qualifiers
1. .625
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="nia:EST:B0970B07-5"
/db_xref="taxon:10090"
/clones="NIA:B0970B07 IMAGE:30478962"
/dev_stage="Unfertilized Egg"
/lab_host="DH10B"
/clone_lib="NIA Mouse Unfertilized Egg cDNA Library (Long 1)"

/note="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI; Site 2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were extracted from a pool of 1488 unfertilized eggs. Double-stranded cDNAs were synthesized with an oligo(dT) primer [Invitrogen: 5'-TGACTAGTCTAGTCGGAGCGGCCCTTTTTTTTTT-3'], treated with T4 DNA polymerase, and purified by

ethanol-precipitation. The cDNAs were ligated to Lone-linker LL-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.5 kb. The library was constructed by Yulan Piao."

ORIGIN

Query Match 100.0%; Score 21; DB 7; Length 625;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGCGC 21
|||||
Db 178 TCAAAGACAACTGTTGCTGCGC 198

RESULT 13

AK009429

LOCUS AK009429

DEFINITION

Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310020017 product:antigenic determinant of rec-A protein, full insert sequence.

ACCESSION AK009429

VERSION AK009429.1 GI:12844218

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (house mouse)

ORGANISM

Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1

Carninci, P. and Hayaehizaki, Y.

AUTHORS High-efficiency full-length cDNA cloning

TITLE Meth. Enzymol. 303, 19-44 (1999)

JOURNAL 10349636

PUBMED

REFERENCE 2

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

AUTHORS Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayaehizaki, Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL Genome Res. 10 (10), 1617-1630 (2000)

PUBMED 11042159

REFERENCE 3

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,

AUTHORS Konno, H., Akiyama, J., Nishi, K., Kitsuunai, T., Tashiro, H., Itoh, M.,

Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,

Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,

Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,

Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayaehizaki, Y.

TITLE RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer

JOURNAL Genome Res. 10 (11), 1757-1771 (2000)

PUBMED 11076861

REFERENCE 4

The RIKEN Genome Exploration Research Group Phase II Team and the

PANTOM Consortium.

TITLE Functional annotation of a full-length mouse cDNA collection

JOURNAL Nature 409, 685-690 (2001)

REFERENCE 5

The PANTOM Consortium and the RIKEN Genome Exploration Research

AUTHORS Group Phase I & II Team.

TITLE Analysis of the mouse transcriptome based on functional annotation

of 60,770 full-length cDNAs

JOURNAL Nature 420, 563-573 (2002)

REFERENCE 6

(bases 1 to 638)

Job time : 120.387 secs

DEFINITION B0917B04-5 NIA Mouse Unfertilized Egg cDNA Library (Long 1) Mus
musculus cDNA: clone NIA:B0917B04 IMAGE:30473871 5', mRNA sequence.
ACCESSION CF172995
VERSION CF172995.1 GI:33282544
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 653)
AUTHORS Piao,Y., Ko,N.T., Lim,M.K. and Ko,M.S.H.
TITLE Construction of long-transcript enriched cDNA libraries from
submicrogram amounts of total RNAs by a universal PCR amplification
method
JOURNAL Genome Res. 11 (9), 1553-1558 (2001)
PUBMED 11544199
COMMENT Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgaun.grc.nia.nih.gov
Plate: B0917 row: B column: 04
Seq primer: M13 Reverse
High quality sequence stop: 653
POLYA=No.
FEATURES
source
1..653
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="niaEST:B0917B04-5"
/db_xref="taxon:10090"
/clone="NIA:B0917B04 IMAGE:30473871"
/dev_stage="Unfertilized Egg"
/lab_host="DH10B"
/clone_lib="NIA Mouse Unfertilized Egg cDNA Library (Long
1)"
/note="vector: pCMV-SPORT6 (Invitrogen); Site_1: SalI;
Site_2: NotI; Mouse cDNA project by the Laboratory of
Genetics, National Institute on Aging (NIA), Intramural
Research Program, NIH (http://lgaun.grc.nia.nih.gov/cDNA).
This is a long-transcript enriched cDNA library (Ref.
Genome Res. 11: 1553-1558 (2001)). [PMID: 11544199]. Total
RNAs were extracted from a pool of 1488 unfertilized eggs.
Double-stranded cDNAs were synthesized with an Oligo(dT)
primer [Invitrogen]:
5'-pGACTAGTCTAGATCGAGCGCGCCCTTTT-3',
treated with T4 DNA polymerase, and purified by
ethanol-precipitation. The cDNAs were ligated to
Lone-linker LL-Sal4, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer Sal4-S. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pCMV-SPORT6 plasmid
vector. The DH10B E. coli host was transformed with the
ligation mixture by the standard chemical method. The
average insert size is about 2.5 kb. The library was
constructed by Yulan Piao."

ORIGIN
Query Match 100.0%; Score 21; DB 6; Length 653;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TCAAAGACAACTGTTGCTGGC 21
Db 168 TCAAAGACAACTGTTGCTGGC 188

Search completed: November 27, 2005, 00:57:58

This Page Blank (uspo)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

QM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:00:02 ; Search time 4.34891 Seconds
(without alignments)
8583.479 Million cell updates/sec

Title: US-09-555-529-27

Perfect score: 21

Sequence: 1 tcaaaqacaactgttctggc 21

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/ptodata/1/ina/1 COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5 COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/6A COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/6B COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/H COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq:*
- 7: /cgn2_6/ptodata/1/ina/PP COMB.seq:*
- 8: /cgn2_6/ptodata/1/ina/RE COMB.seq:*
- 9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17.8	84.8	601	3	US-09-949-002-8571
2	17.8	84.8	601	3	US-09-949-002-8572
3	17.8	84.8	601	3	US-09-949-002-8573
4	17.8	84.8	601	3	US-09-949-002-8574
5	17.8	84.8	26191	3	US-09-949-002-8574
6	17.8	84.8	39428	3	US-09-949-002-8574
7	16.8	80.0	780	3	US-09-328-352-928
8	16.8	80.0	133559	3	US-09-949-016-15845
9	16.2	77.1	384	3	US-09-902-540-2124
10	16.2	77.1	601	3	US-09-949-016-29076
11	16.2	77.1	601	3	US-09-949-016-69609
12	16.2	77.1	617	3	US-09-902-540-1335
13	16.2	77.1	1946	3	US-08-961-527-239
14	16.2	77.1	1075	3	US-09-949-016-3994
15	16.2	77.1	1854	3	US-09-107-433-1993
16	16.2	77.1	1977	3	US-09-583-110-2082
17	16.2	77.1	1998	3	US-09-543-681A-1706
18	16.2	77.1	17328	3	US-09-949-016-15736
19	16.2	77.1	20537	3	US-09-949-016-12887
20	16.2	77.1	20538	3	US-09-949-016-17052
21	16.2	77.1	34628	3	US-09-949-016-12304
22	16.2	77.1	34779	3	US-09-949-016-13787
23	16.2	77.1	145287	3	US-09-949-016-13530
24	16.2	77.1	145287	3	US-09-949-016-13531

C	25	16	76.2	397	3	US-09-513-999C-32957	Sequence 32957, A
	26	15.8	75.2	519	3	US-08-669-408A-7	Sequence 7, Appl1
	27	15.8	75.2	529	3	US-09-621-976-16620	Sequence 16620, A
	28	15.8	75.2	549	3	US-09-621-976-992	Sequence 992, App
	29	15.8	75.2	601	3	US-09-949-016-13880	Sequence 13880, App
	30	15.8	75.2	601	3	US-09-949-016-140951	Sequence 140951, App
	31	15.8	75.2	687	3	US-09-248-796A-3585	Sequence 3585, App
	32	15.8	75.2	822	3	US-09-792-024-12	Sequence 12, Appl1
	33	15.8	75.2	1429	3	US-09-516-143A-5	Sequence 5, Appl1
	34	15.8	75.2	1429	3	US-09-984-205-5	Sequence 5, Appl1
	35	15.8	75.2	1429	3	US-10-759-277-5	Sequence 1, Appl1
	36	15.8	75.2	1429	3	US-09-930-440C-1	Sequence 1, Appl1
	37	15.8	75.2	1749	3	US-09-248-796A-5536	Sequence 5536, App
	38	15.8	75.2	2536	2	US-09-027-013-2	Sequence 2, Appl1
	39	15.8	75.2	2536	3	US-09-244-233-2	Sequence 2, Appl1
	40	15.8	75.2	5092	3	US-09-620-312D-153	Sequence 153, App
	41	15.8	75.2	31459	3	US-09-949-016-13722	Sequence 13722, A
	42	15.8	75.2	37585	3	US-09-949-016-15847	Sequence 15847, A
	43	15.8	75.2	42571	3	US-09-810-347-3	Sequence 3, Appl1
	44	15.8	75.2	58782	3	US-09-949-016-16851	Sequence 16851, A
	45	15.8	75.2	60785	3	US-09-949-016-12774	Sequence 12774, A

ALIGNMENTS

RESULT 1

US-09-949-002-8571
; Sequence 8571, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8571
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; US-09-949-002-8571

Query Match 84.8%; Score 17.8; DB 3; Length 601;
Best Local Similarity 90.5%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 TCAAAGACAACACTGTTGCTGGC 21
|||
Db 44 TCAAAGACAACACTGTTGCTGGC 64

RESULT 2

US-09-949-002-8572
; Sequence 8572, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0

```
; SEQ ID NO 8572
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-8572

Query Match      84.8%; Score 17.8; DB 3; Length 601;
Best Local Similarity 90.5%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGGC 21
    |||||
Db 245 TCAAAGACAACTGATGTTGGC 265

RESULT 3
US-09-949-002-8573
; Sequence 8573, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8573
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-8573

Query Match      84.8%; Score 17.8; DB 3; Length 601;
Best Local Similarity 90.5%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGGC 21
    |||||
Db 350 TCAAAGACAACTGATGTTGGC 370

RESULT 4
US-09-949-002-8574
; Sequence 8574, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8574
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-8574

Query Match      84.8%; Score 17.8; DB 3; Length 601;
Best Local Similarity 90.5%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGGC 21
    |||||
Db 1 TCAAAGACAACTGATGTTGGC 21

RESULT 5
US-09-949-002-686/c
; Sequence 686, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 686
; LENGTH: 26191
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-686

Query Match      84.8%; Score 17.8; DB 3; Length 26191;
Best Local Similarity 90.5%; Pred. No. 48;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGGC 21
    |||||
Db 9629 TCAAAGACAACTGATGTTGGC 9609

RESULT 6
US-09-949-002-794/c
; Sequence 794, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 794
; LENGTH: 39428
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-794

Query Match      84.8%; Score 17.8; DB 3; Length 39428;
Best Local Similarity 90.5%; Pred. No. 52;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGGC 21
    |||||
Db 7638 TCAAAGACAACTGATGTTGGC 7618

RESULT 7
US-09-328-352-928
; Sequence 928, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
```



```
,
, NUMBER OF SEQ ID NOS: 207012
,
, FASTQ: FastSeq for Windows Version 4.0
,
, SEQ ID NO 69609
,
, LENGTH: 601
,
, TYPE: DNA
,
, ORGANISM: Human
,
```

US-09-949-016-69609

Query Match 77.1%; Score 16.2; DB 3; Length 601;
Best Local Similarity 85.7%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGGC 21
||||| ||||| ||||| ||||| |||||
DB 81 TCAAAGCCCAACTGTTGTTGAC 61

RESULT 12

US-09-902-540-1335/c
Sequence 1335, Application US/09902540
Patent No. 6833447

GENERAL INFORMATION:
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10(15849)B
CURRENT APPLICATION NUMBER: US/09/902,540
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/217,883
PRIOR FILING DATE: 2000-07-10
NUMBER OF SEQ ID NOS: 16825
SEQ ID NO 1335
LENGTH: 617
TYPE: DNA
ORGANISM: Myxococcus xanthus

US-09-902-540-1335

Query Match 77.1%; Score 16.2; DB 3; Length 617;
Best Local Similarity 85.7%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGGC 21
||||| ||||| ||||| ||||| |||||
DB 147 TCAAACCAACTGTTGCTAGC 127

RESULT 13

US-08-961-527-239/c
Sequence 239, Application US/08961527
Patent No. 6420135

GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 239:
SEQUENCE CHARACTERISTICS:
LENGTH: 946 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-239

Query Match 77.1%; Score 16.2; DB 3; Length 946;
Best Local Similarity 85.7%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGGC 21
||||| ||||| ||||| ||||| |||||
DB 617 TCAAAGACATCTGTAGATGGC 597

RESULT 14

US-09-949-016-3994
Sequence 3994, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3994
LENGTH: 1075
TYPE: DNA
ORGANISM: Human
US-09-949-016-3994

Query Match 77.1%; Score 16.2; DB 3; Length 1075;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGGC 21
||||| ||||| ||||| ||||| |||||
DB 712 TCACAGACATCTGTTGCTGCC 732

RESULT 15

US-09-107-433-1993/c
Sequence 1993, Application US/09107433
Patent No. 6800744

GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE THERAPEUTICS
NUMBER OF SEQUENCES: 5206
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM: CD-ROM ISO9660
MEDIUM TYPE: CD-ROM ISO9660

FOR DIAGNOSIS

```

;
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107.433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
;
; INFORMATION FOR SEQ ID NO: 1993:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1854 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1854
; SEQUENCE DESCRIPTION: SEQ ID NO: 1993:
US-09-107-433-1993
;
; Query Match
; Best Local Similarity 85.7%; Score 16.2; DB 3; Length 1854;
; Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
;
; Qy 1 TCAAAGACAACTGTTGCTGGC 21
; Db 1631 TCAAAGACATCTGTAGATGGC 1611

```

Search completed: November 27, 2005, 01:10:50
Job time : 5.34891 secs

This Page Blank (uspo1c)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:00:41 ; Search time 24.2095 Seconds
(without alignments)
7173.088 Million cell updates/sec

Title: US-09-555-529-27

Perfect score: 21

Sequence: 1 tcaagacacactgttctgctgac 21

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.Main:

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.4	87.6	601	5	US-10-027-632-6241
2	18.4	87.6	601	6	US-10-027-632-6241
3	18.4	87.6	1389	7	US-10-437-963-50742
4	18.4	87.6	2250	7	US-10-437-963-50697
5	18.4	87.6	2692	7	US-10-437-963-50728
6	18.4	87.6	2754	7	US-10-437-963-50685
7	18.4	87.6	3735	7	US-10-437-963-86964
8	18.4	87.6	3831	7	US-10-437-963-86972
9	18.4	87.6	3897	7	US-10-437-963-50771
10	18.4	87.6	3930	7	US-10-437-963-86963
11	18.4	87.6	4275	7	US-10-437-963-50693
12	18.4	87.6	4347	7	US-10-437-963-50692
13	18.4	87.6	4353	7	US-10-437-963-50691
14	18.4	87.6	4497	7	US-10-437-963-50727
15	18.4	87.6	4506	7	US-10-437-963-50704
16	18.4	87.6	4518	7	US-10-437-963-50778
17	18.4	87.6	4551	7	US-10-437-963-50696
18	18.4	87.6	4551	7	US-10-437-963-50738
19	18.4	87.6	4710	7	US-10-437-963-50734
20	18.4	87.6	4843	7	US-10-437-963-50780
21	18.4	87.6	4881	7	US-10-437-963-86965
22	18.4	87.6	4902	7	US-10-437-963-50772
23	18.4	87.6	4917	7	US-10-437-963-50774

C 24	18.4	87.6	10319	3	US-09-764-868-1292	Sequence 1292, A
C 25	17.8	84.8	1061	5	US-10-027-632-30377	Sequence 30377, A
C 26	17.8	84.8	1061	5	US-10-027-632-30378	Sequence 30378, A
C 27	17.8	84.8	1061	6	US-10-027-632-30377	Sequence 30377, A
C 28	17.8	84.8	1061	6	US-10-027-632-30378	Sequence 30378, A
C 29	17.8	84.8	2454	7	US-10-437-963-35868	Sequence 35868, A
C 30	17.8	84.8	3040	7	US-10-437-963-47595	Sequence 47595, A
C 31	17.8	84.8	12555	3	US-09-764-847-1187	Sequence 1187, A
C 32	17.8	84.8	12555	5	US-10-092-154-1187	Sequence 1187, A
C 33	17.8	84.8	13810	10	US-11-097-143-6271	Sequence 6271, A
C 34	17	81.0	497	7	US-10-424-599-74853	Sequence 74853, A
C 35	17	81.0	1793	7	US-10-424-599-74852	Sequence 74852, A
C 36	16.8	80.0	306	8	US-10-425-115-21694	Sequence 21694, A
C 37	16.8	80.0	461	7	US-10-242-535A-29177	Sequence 29177, A
C 38	16.8	80.0	461	7	US-10-085-783A-29177	Sequence 29177, A
C 39	16.8	80.0	565	4	US-09-925-065A-262385	Sequence 262385, A
C 40	16.8	80.0	600	9	US-10-972-079-15579	Sequence 15579, A
C 41	16.8	80.0	600	9	US-10-972-079-96162	Sequence 96162, A
C 42	16.8	80.0	684	8	US-10-425-115-85493	Sequence 85493, A
C 43	16.8	80.0	1305	7	US-10-437-963-14622	Sequence 14622, A
C 44	16.8	80.0	1405	8	US-10-425-115-12565	Sequence 12565, A
C 45	16.8	80.0	1702	4	US-09-925-065A-56826	Sequence 56826, A

ALIGNMENTS

RESULT 1
US-10-027-632-6241
; Sequence 6241, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6241
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(601)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-6241

Query Match 87.6%; Score 18.4; DB 5; Length 601;
Best Local Similarity 95.0%; Pred. No. 63;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CAAGACAACTGTGCTGGC 21

Db 109 CAAGACAACTGTGCTGGC 128

```
RESULT 2
US-10-027-632-6241
; Sequence 6241, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6241
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(601)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-6241
```

```
Query Match      87.6%; Score 18.4; DB 6; Length 601;
Best Local Similarity 95.0%; Pred. No. 63;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      2  CAAAGACAACCTGTTGCTGGC 21
Db      109  CAAAGACAACCTGTTGCTGGC 128
```

```
RESULT 3
US-10-437-963-50742/c
; Sequence 50742, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; File of Invention: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 50742
; LENGTH: 1389
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_53199C.1
US-10-437-963-50742
```

```
Query Match      87.6%; Score 18.4; DB 7; Length 1389;
Best Local Similarity 95.0%; Pred. No. 74;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1  TCAAAGACAACCTGTTGCTGG 20
Db      1154  TCAAAGAGAACTGTTGCTGG 1135

RESULT 4
US-10-437-963-50697/c
; Sequence 50697, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; File of Invention: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 50697
; LENGTH: 2250
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_53158C.1
US-10-437-963-50697
```

```
Query Match      87.6%; Score 18.4; DB 7; Length 2250;
Best Local Similarity 95.0%; Pred. No. 80;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1  TCAAAGACAACCTGTTGCTGG 20
Db      1352  TCAAAGAGAACTGTTGCTGG 1333
```

```
RESULT 5
US-10-437-963-50728/c
; Sequence 50728, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; File of Invention: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 50728
; LENGTH: 2692
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_53186C.1
US-10-437-963-50728
```

```
Query Match      87.6%; Score 18.4; DB 7; Length 2692;
```

Best Local Similarity 95.0%; Pred. No. 83;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGG 20
|||||
Db 661 TCAAAGAGAACTGTTGCTGG 642

RESULT 6

US-10-437-963-50685/c
; Sequence 50685, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 50685
; LENGTH: 2754
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_53147C.1
US-10-437-963-50685

Query Match 87.6%; Score 18.4; DB 7; Length 2754;
Best Local Similarity 95.0%; Pred. No. 83;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGG 20
|||||
Db 1226 TCAAAGAGAACTGTTGCTGG 1207

RESULT 7

US-10-437-963-86964/c
; Sequence 86964, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 86964
; LENGTH: 3735
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_85957C.1
US-10-437-963-86964

Query Match 87.6%; Score 18.4; DB 7; Length 3735;
Best Local Similarity 95.0%; Pred. No. 88;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TCAAAGACAACTGTTGCTGG 20
|||||
Db 932 TCAAAGAGAACTGTTGCTGG 913

RESULT 8

US-10-437-963-86972/c
; Sequence 86972, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 86972
; LENGTH: 3831
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_85964C.1
US-10-437-963-86972

Query Match 87.6%; Score 18.4; DB 7; Length 3831;
Best Local Similarity 95.0%; Pred. No. 89;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGG 20
|||||
Db 1088 TCAAAGAGAACTGTTGCTGG 1069

RESULT 9

US-10-437-963-50771/c
; Sequence 50771, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 50771
; LENGTH: 3897
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_53224C.1
US-10-437-963-50771

Query Match 87.6%; Score 18.4; DB 7; Length 3897;
Best Local Similarity 95.0%; Pred. No. 89;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCAAAGACAACACTGTTGCTGG 20
|||||
DB 2999 TCAAAGACAACACTGTTGCTGG 2980

RESULT 10
US-10-437-963-86963/c
; Sequence 86963, Application US/10437963
; Publication NO. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 86963
; LENGTH: 3930
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_85956C.1
US-10-437-963-86963

Query Match 87.6%; Score 18.4; DB 7; Length 3930;
Best Local Similarity 95.0%; Pred. No. 89;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCAAAGACAACACTGTTGCTGG 20
|||||
DB 875 TCAAAGACAACACTGTTGCTGG 856

RESULT 11
US-10-437-963-50693/c
; Sequence 50693, Application US/10437963
; Publication NO. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 50693
; LENGTH: 4275
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_53154C.1
US-10-437-963-50693

Query Match 87.6%; Score 18.4; DB 7; Length 4275;
Best Local Similarity 95.0%; Pred. No. 90;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCAAAGACAACACTGTTGCTGG 20
|||||
DB 1157 TCAAAGACAACACTGTTGCTGG 1138

RESULT 12
US-10-437-963-50692/c
; Sequence 50692, Application US/10437963
; Publication NO. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 50692
; LENGTH: 4347
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_53153C.1
US-10-437-963-50692

Query Match 87.6%; Score 18.4; DB 7; Length 4347;
Best Local Similarity 95.0%; Pred. No. 91;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCAAAGACAACACTGTTGCTGG 20
|||||
DB 1226 TCAAAGACAACACTGTTGCTGG 1207

RESULT 13
US-10-437-963-50691/c
; Sequence 50691, Application US/10437963
; Publication NO. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 50691
; LENGTH: 4353
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_53152C.1
US-10-437-963-50691

Query Match 87.6%; Score 18.4; DB 7; Length 4353;
Best Local Similarity 95.0%; Pred. No. 91;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCAAAGACAACACTGTTGCTGG 20


```
Db      1226 TCAAAGAGAACTGTTGCTGG 1207
||||||| ||||| ||||| |||||
1181 TCAAAGAGAACTGTTGCTGG 1162

Search completed: November 27, 2005, 02:22:43
Job time : 25.2095 secs

RESULT 14
US-10-437-963-50727/c
; Sequence 50727, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 50727
; LENGTH: 4497
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_53185C.1
US-10-437-963-50727

Query Match      87.6%; Score 18.4; DB 7; Length 4497;
Best Local Similarity 95.0%; Pred. No. 91;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TCAAAGAGAACTGTTGCTGG 20
||||||| ||||| ||||| |||||
Db      1226 TCAAAGAGAACTGTTGCTGG 1207

RESULT 15
US-10-437-963-50704/c
; Sequence 50704, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 50704
; LENGTH: 4506
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_53164C.1
US-10-437-963-50704

Query Match      87.6%; Score 18.4; DB 7; Length 4506;
Best Local Similarity 95.0%; Pred. No. 91;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TCAAAGAGAACTGTTGCTGG 20
||||||| ||||| ||||| |||||
```

This Page Blank (uspo10)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:41:08 ; Search time 7.41177 Seconds
(without alignments)
421.056 Million cell updates/sec

Title: US-09-555-529-27

Perfect score: 21 tcaagacaaactgttctgtgc 2i

Sequence: 1 tcaagacaaactgttctgtgc 2i

Scoring table: IDENTITY INUC

Gapop 10% , Gapext 1.0

Searched: 3205263 seqs, 74304013 residues

Total number of hits satisfying chosen parameters: 6410526

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA New:*

- 1: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	15.2	72.4	903	1	US-10-793-626-1899
C 2	15.2	72.4	929	1	Sequence 1899, Ap
C 3	15.2	72.4	2997	1	Sequence 1311, Ap
C 4	15.2	72.4	3269	1	Sequence 3723, Ap
C 5	15.2	72.4	3314	1	Sequence 3692, Ap
C 6	14.8	70.5	19	8	Sequence 3927, Ap
C 7	14.8	70.5	19	9	Sequence 1021845,
C 8	14.8	70.5	876	1	Sequence 1021845,
C 9	14.8	70.5	1705	1	Sequence 915, App
C 10	14.8	70.5	3532	1	Sequence 115, App
C 11	14.8	70.5	3657	1	Sequence 4153, Ap
C 12	14.8	70.5	4010	1	Sequence 3888, Ap
C 13	14.4	68.6	19	8	Sequence 4296, Ap
C 14	14.4	68.6	19	8	Sequence 146179,
C 15	14.4	68.6	19	8	Sequence 146284,
C 16	14.4	68.6	19	9	Sequence 146179,
C 17	14.4	68.6	19	9	Sequence 146284,
C 18	14.4	68.6	19	9	Sequence 146179,
C 19	14.2	67.6	19	8	Sequence 146284,
C 20	14.2	67.6	19	8	Sequence 146179,
C 21	14.2	67.6	19	9	Sequence 146284,
C 22	14.2	67.6	19	9	Sequence 146179,
C 23	14.2	67.6	447	1	Sequence 18, Appl

24	14.2	67.6	839	1	US-10-512-109-4	Sequence 4, Appli
25	14.2	67.6	2754	1	US-10-821-234-163	Sequence 163, App
26	14.2	67.6	40439	1	US-10-993-509-1	Sequence 1, Appli
27	14.2	67.6	48000	7	US-11-159-597-20	Sequence 20, Appl
C 28	14.2	67.6	55763	1	US-10-972-766-1	Sequence 1, Appli
C 29	13.8	65.7	19	8	US-11-101-244-154358	Sequence 154358,
C 30	13.8	65.7	19	8	US-11-101-244-154405	Sequence 154405,
C 31	13.8	65.7	19	8	US-11-101-244-1437747	Sequence 1437747,
C 32	13.8	65.7	19	9	US-11-083-784-154358	Sequence 154358,
C 33	13.8	65.7	19	9	US-11-083-784-154405	Sequence 154405,
C 34	13.8	65.7	19	9	US-11-083-784-1437747	Sequence 1437747,
C 35	13.8	65.7	885	1	US-10-858-730-188	Sequence 188, App
C 36	13.8	65.7	2598	1	US-10-467-962B-32	Sequence 32, Appl
C 37	13.8	65.7	3048	1	US-10-957-569-50	Sequence 50, Appl
C 38	13.8	65.7	17410	7	US-11-051-568-3	Sequence 3, Appli
C 39	13.6	64.8	56	7	US-11-013-533-5	Sequence 5, Appli
C 40	13.6	64.8	1293	7	US-11-074-176-131	Sequence 131, App
C 41	13.6	64.8	1420	1	US-10-512-109-5	Sequence 5, Appli
C 42	13.6	64.8	1722	1	US-10-510-386-113	Sequence 113, App
C 43	13.6	64.8	3132	1	US-10-512-109-22	Sequence 22, Appl
C 44	13.6	64.8	5796	1	US-10-821-234-62	Sequence 62, Appl
C 45	13.6	64.8	7512	1	US-10-647-956A-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-10-793-626-1899/c
; Sequence 1899, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1899
; LENGTH: 903
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-10-793-626-1899

Query Match	72.4%	Score 15.2;	DB 1;	Length 903;
Best Local Similarity	85.0%	Pred. No. 51;		
Mismatches	17;	Conservative	0;	Mismatches
			3;	Indels
				Gaps
				0;

Qy	2	CACAGACAACTGTTGCTGGC	21
Db	118	CACGATAACTGTTGCTGGC	99

RESULT 2
US-10-793-626-1311
; Sequence 1311, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1311
; LENGTH: 929
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-1311

Query Match 72.4%; Score 15.2; DB 1; Length 929;
Best Local Similarity 85.0%; Pred. No. 51;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGG 20
DB 754 TCTAAACAACTGTTACTGG 773

RESULT 3
US-10-793-626-3723/c
; Sequence 3723, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3723
; LENGTH: 2997
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3723

Query Match 72.4%; Score 15.2; DB 1; Length 2997;
Best Local Similarity 85.0%; Pred. No. 61;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGG 20
DB 176 TCTAAACAACTGTTACTGG 157

RESULT 4
US-10-793-626-3692
; Sequence 3692, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3692
; LENGTH: 3269
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3692

Query Match 72.4%; Score 15.2; DB 1; Length 3269;
Best Local Similarity 85.0%; Pred. No. 62;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CAAAGACAACTGTTGCTGGC 21
DB 2986 CACCGATAACTGTTGCTGGC 3005

RESULT 5
US-10-793-626-3927
; Sequence 3927, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3927
; LENGTH: 3314
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3927

Query Match 72.4%; Score 15.2; DB 1; Length 3314;
Best Local Similarity 85.0%; Pred. No. 62;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CAAAGACAACTGTTGCTGGC 21
DB 959 CACCGATAACTGTTGCTGGC 978

RESULT 6
US-11-101-244-1021845/c
; Sequence 1021845, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2003-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 1021845
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-1021845

Query Match 70.5%; Score 14.8; DB 8; Length 19;
Best Local Similarity 88.9%; Pred. No. 46;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CAAAGACAACTGTTGCTG 19

```
Db      19 CAAAGAAACTGTTGTTG 2
||||| ||||| ||||| |||
RESULT 7
US-11-083-784-1021845/c
; Sequence 1021845, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 134980S
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR FILING DATE: US/10/714,333
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 1021845
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-1021845
Query Match      70.5%; Score 14.8; DB 9; Length 19;
Best Local Similarity 88.9%; Pred. No. 46;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2 CAAAGACAACTGTTGCTG 19
||||| ||||| ||||| |||
Db      19 CAAAGAAACTGTTGTTG 2

RESULT 8
US-10-793-626-915
; Sequence 915, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 915
; LENGTH: 876
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-915
Query Match      70.5%; Score 14.8; DB 1; Length 876;
Best Local Similarity 88.9%; Pred. No. 81;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2 CAAAGACAACTGTTGCTG 19
||||| ||||| ||||| |||
Db      806 CAAAGACTATTGTTGCTG 823

RESULT 9
US-10-510-386-115/c
; Sequence 115, Application US/10510386
; Publication No. US2005024922A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Jens Tonne
; APPLICANT: Clausen, Ib Groth
; APPLICANT: Jorgensen, Steen Troels
; APPLICANT: Olsen, Peter Bjarke
; APPLICANT: Rasmussen, Michael Dolberg
; TITLE OF INVENTION: Improved Bacillus Host Cell
; FILE REFERENCE: 10294.204-US
; CURRENT APPLICATION NUMBER: US/10/510,386
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 248
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 115
; LENGTH: 1705
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (507)..(1202)
US-10-510-386-115
Query Match      70.5%; Score 14.8; DB 1; Length 1705;
Best Local Similarity 88.9%; Pred. No. 89;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      4 AAGACAACTGTTGCTGCG 21
||||| ||||| ||||| |||
Db      1311 AAGCCAACTGTTGCTGCG 1294

RESULT 10
US-10-793-626-4153
; Sequence 4153, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4153
; LENGTH: 3532
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4153
Query Match      70.5%; Score 14.8; DB 1; Length 3532;
Best Local Similarity 88.9%; Pred. No. 99;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2 CAAAGACAACTGTTGCTG 19
||||| ||||| ||||| |||
Db      2808 CAAAGACTATTGTTGCTG 2825

RESULT 11
US-10-793-626-3888
; Sequence 3888, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
```

```
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3888
; LENGTH: 3657
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3888

Query Match          70.5%; Score 14.8; DB 1; Length 3657;
Best Local Similarity 88.9%; Pred. No. 99;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 CAAAGACAACACTGTTGCTG 19
       ||||| ||||| |||||
Db      789 CAAAGACTATTGTTGCTG 806

RESULT 12
US-10-793-626-4296/c
; Sequence 4296, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4296
; LENGTH: 4010
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4296

Query Match          70.5%; Score 14.8; DB 1; Length 4010;
Best Local Similarity 88.9%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 CAAAGACAACACTGTTGCTG 19
       ||||| ||||| |||||
Db      2242 CAAAGACTATTGTTGCTG 2225

RESULT 13
US-11-101-244-146179/c
; Sequence 146179, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
```

```
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 146179
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-146179

Query Match          68.6%; Score 14.4; DB 8; Length 19;
Best Local Similarity 93.8%; Pred. No. 73;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TCAAAGACAACACTGTTG 16
       ||||| ||||| |||||
Db      18 TCAAAGACAGCTGTTG 3

RESULT 14
US-11-101-244-146284/c
; Sequence 146284, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 146284
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-146284

Query Match          68.6%; Score 14.4; DB 8; Length 19;
Best Local Similarity 93.8%; Pred. No. 73;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TCAAAGACAACACTGTTG 16
       ||||| ||||| |||||
Db      18 TCAAAGACAGCTGTTG 3

RESULT 15
US-11-101-244-146377/c
; Sequence 146377, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
```

; PRIOR FILING DATE: 2003-09-10
 ; PRIOR APPLICATION NUMBER: 60/426,137
 ; PRIOR FILING DATE: 2002-11-14
 ; NUMBER OF SEQ ID NOS: 1591911
 ; SOFTWARE: Proprietary
 ; SEQ ID NO 146377
 ; LENGTH: 19
 ; TYPE: RNA
 ; ORGANISM: Homo sapiens
 US-11-101-244-146377

Query Match : 68.6%; Score 14.4; DB 8; Length 19;
 Best Local Similarity : 93.8%; Pred. No. 73;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTG 16
 |||||
 Db 18 TCAAAGACAGCTGTTG 3

Search completed: November 27, 2005, 02:44:31
 Job time : 7.41177 secs

This Page Blank (uspi.c)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 12:53:47 ; Search time 116.932 Seconds
(without alignments)
10694.773 Million cell updates/sec

Title: US-09-555-529-28

Perfect score: 22

Sequence: 1 atacctcaactctgcgtcctt 22

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_in.*

3: gb_env.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pr.*

9: gb_ro.*

10: gb_sts.*

11: gb_sy.*

12: gb_un.*

13: gb_vi.*

14: gb_htg.*

15: gb_pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	100.0	22	6	AX003335 Sequence
C 2	22	100.0	333	6	AX003341 Sequence
C 3	22	100.0	1102	6	AX003309 Sequence
C 4	22	100.0	1386	9	BC028860 Mus muscu
C 5	22	100.0	1390	6	AX003331 Sequence
C 6	22	100.0	1414	9	WVKIN17
C 7	22	100.0	2710	9	BC058169 Mus muscu
C 8	22	100.0	180882	14	AC114615
C 9	22	100.0	182400	14	AC124010 Mus muscu
C 10	22	100.0	185378	9	AL772367
C 11	22	100.0	221547	14	AC119716
C 12	20.4	92.7	22	6	AX003328 Sequence
C 13	20.4	92.7	333	6	AX003340 Sequence
C 14	20.4	92.7	1002	6	AX003310 Sequence
C 15	20.4	92.7	1002	6	AX003311 Sequence
C 16	20.4	92.7	1528	6	AX003308 Sequence
C 17	20.4	92.7	1528	8	HSBJ5273
C 18	20.4	92.7	2069	8	BC017309 Homo sapi

AL158044 Human DNA
AL391687 Homo sapi
AR547015 Sequence
273248 Bovine vira
AF065434 Candida a
AF002227 Bordar di
AY149215 Bovine vi
AY149216 Bovine vi
AX741975 Sequence
AF502399 Bovine vi
AF145967 Bovine vi
AC012951 Drosophil
Continuation (2 of
AC011995 Rattus no
AC003374 Drosophil
AC097562 Rattus no
AC160550 Mus muscu
BX842664 Mouse DNA
AL844903 Mouse DNA
AC157305 Bos tauru
AC094962 Rattus no
AY442521 Rattus vi
U18059 Bovine vira
BX248133 Human DNA
AC149956 Strongylo
CR925765 Human DNA
AC100656 Mus muscu

ALIGNMENTS

RESULT 1
AX003335
LOCUS AX003335 22 bp DNA linear PAT 24-AUG-2000
DEFINITION Sequence 28 from Patent WO9929845.
ACCESSION AX003335
VERSION AX003335.1 GI:9927148

SOURCE Mus sp.

ORGANISM Mus sp.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE 1 Angulo-Mora, J.F. and Mauffrey, P.

AUTHORS Sequences coding for kin17 protein and their applications

TITLE Patent: WO 9929845-A 28 17-JUN-1999;

JOURNAL ANGULO MORA JAIME FRANCISCO (FR) ; COMMISSARIAT ENERGIE ATOMIQUE

(FR)

FEATURES Location/Qualifiers

source
1..22
/organism="Mus sp."
/mol_type="unassigned DNA"
/db_xref="taxon:10095"

ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 22;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATACCTTCAACTCTGCGTCCTT 22

Db 1 ATACCTTCAACTCTGCGTCCTT 22

RESULT 2

AX003341/c

LOCUS AX003341 333 bp DNA linear PAT 24-AUG-2000

DEFINITION Sequence 34 from Patent WO9929845.

ACCESSION AX003341

VERSION AX003341.1 GI:9927154

KEYWORDS

/protein_id="AAH28860.1"
 /db_xref="GI:22135638"
 /db_xref="GeneID:16588"
 /db_xref="MGI:96676"
 /translation="MGKSDFLSPKAIANRIKSKGLQKLRYWCOMQKQCRDENGFKCH
 CSESHQROQLLASENPQOFMDYFSEFRNDFLELLRRFGTKRVHNNIVYNYISHR
 EHTHNAQTWETLDTFKLGRGLCKVDTPKGYIOYIDRDPETIRQLEKKKK
 ODLDDEKTAFTFEORVRLGEGKEOTVFTFELSRENEEKVTFNLNKAGGAGCAT
 TSKSSLSGSPALKLLSASAGKRKSSQSSAPAKKKKALDIMELEBKRTATD
 AWLPQGLVKKIITKKLGKHKKGVVKEVIDRYTAVRWTDSDRLKLDQTHLETVI
 PAPGKRVLVNGGYRGNEGTLSEINEKAFSATIVETGPKLGRRVGEGIQVEDISKLA"

ORIGIN

Query Match 100.0%; Score 22; DB 9; Length 1386;
 Best Local Similarity 100.0%; Pred. No. 2.8;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATACCTTCAACTCTGCGTCCTT 22
 |||||
 Db 1164 ATACCTTCAACTCTGCGTCCTT 1143

RESULT 5

AX003331/c
 LOCUS 1390 bp DNA linear PAT 24-AUG-2000
 DEFINITION Sequence 24 from Patent WO9929845.
 ACCESSION AX003331
 VERSION AX003331.1 GI:9927146
 KEYWORDS
 SOURCE Mus sp.
 ORGANISM Mus sp.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE

1
 AUTHORS Angulo-Mora, J. F. and Mauffrey, P.
 TITLES Sequences coding for kin17 protein and their applications
 JOURNAL Patent: WO 9929845-A 24 17-JUN-1999;
 ANGULO MORA JAINE FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE (FR)

FEATURES

source
 1..1390
 /organism="Mus sp."
 /mol_type="unassigned DNA"
 /db_xref="taxon:10095"

ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 1390;
 Best Local Similarity 100.0%; Pred. No. 2.8;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATACCTTCAACTCTGCGTCCTT 22
 |||||
 Db 1145 ATACCTTCAACTCTGCGTCCTT 1124

RESULT 6

MMKIN17/c
 LOCUS 1414 bp mRNA linear ROD 30-OCT-1995
 DEFINITION Mouse KIN17 mRNA for kin17 protein.
 ACCESSION X58472
 VERSION X58472.1 GI:1045209
 KEYWORDS KIN17 gene; kin17 protein; zinc-finger protein.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 1414)

Auteurs Angulo, J. F., Rouer, E., Benarous, R. and Devoret, R.
 TITLES Identification of a mouse cDNA fragment whose expressed polypeptide
 JOURNAL reacts with anti-recA antibodies
 Biochimie 73 (2-3), 251-256 (1991)

1715759

REFERENCE

2
 AUTHORS Angulo, J. F., Rouer, E., Mazin, A., Mattei, M. G., Tissier, A.,
 Horellou, P., Benarous, R. and Devoret, R.
 TITLES Identification and expression of the cDNA of KIN17, a zinc-finger
 gene located on mouse chromosome 2, encoding a new DNA-binding
 protein

Nucleic Acids Res. 19 (19), 5117-5123 (1991)

1923796

REFERENCE

3

Angulo, J.

Direct Submission

Submitted (25-MAR-1991) J. Angulo, C N R S, GEMC-Lab d'Enzymologie,

1 Avenue de la Terrasse, Gif-Sur-Yvette, Ile-de-France, FRANCE

Revised by (4) MGPp

4 (bases 1 to 1414)

Angulo, J.

Direct Submission

Submitted (30-OCT-1995) J. Angulo, C N R S, GEMC-Lab d'Enzymologie,

1 Avenue de la Terrasse, Gif-Sur-Yvette, Ile-de-France, FRANCE

On Oct 31, 1995 this sequence version replaced gi:52795.

FEATURES

Location/Qualifiers

1..1414

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/cell_line="MB66"

/tissue_type="brain"

/dev_stage="embryo"

1..1414

/gene="KIN17"

/pseudo

<1..1382

/gene="KIN17"

25..1200

/gene="KIN17"

/pseudo

/codon_start=1

/product="kin17 protein"

/db_xref="PSEUDO:CAA41386.1"

106..174

/gene="KIN17"

/notes="zinc-finger of kin17"

/pseudo

234..842

/gene="KIN17"

/notes="kin17(200) fragment"

/pseudo

/evidence=experimental

472..496

/gene="KIN17"

/notes="kin17 helix1 motif"

/pseudo

781..792

/gene="KIN17"

/notes="nuclear localization signal 1"

/pseudo

907..924

/gene="KIN17"

/notes="nuclear localization signal 2"

/pseudo

1382

/gene="KIN17"

/pseudo

1382

/gene="KIN17"

/pseudo

1382

/gene="KIN17"

ORIGIN

Query Match 100.0%; Score 22; DB 9; Length 1414;

Best Local Similarity 100.0%; Pred. No. 2.8;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATACCTTCAACTCTGCGTCCTT 22

|||||

Db 1169 ATACCTTCAACTCTGCGTCCTT 1148

Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Menuez, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (09-MAR-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 2, 2004 this sequence version replaced gi:24182248.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L24315
Center clone name: 82_I_19
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 191018 bases at least Q40
Consensus quality: 191303 bases at least Q30
Consensus quality: 191388 bases at least Q20
Insert size: 178000; agarose-fp
Insert size: 191574; sum-of-contigs
Quality coverage: 12.2 in Q20 bases; agarose-fp
Quality coverage: 11.3 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 89963: contig of 89963 bp in length
* 89964 90063: gap of 100 bp
* 90064 94560: contig of 4497 bp in length
* 94561 94660: gap of 100 bp
* 94661 169601: contig of 74941 bp in length
* 169602 169701: gap of 100 bp
* 169702 180882: contig of 11181 bp in length.

FEATURES
source
1. .180882
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP24-82119"
/clone_lib="RPC1-24 Male Mouse BAC"
1. .89963
/note="assembly_fragment
clone end:SP6
vector side:left"
89964..90063
/estimated_length=100
90064..94560
/note="assembly_fragment"
94561..94660
/estimated_length=100
94661..169601

misc_feature
1. .89963
/note="assembly_fragment
clone end:T7
vector side:right"

gap
misc_feature
169602..169701
/estimated_length=100
169702..180882
/note="assembly_fragment
clone end:T7
vector side:right"

ORIGIN
Query Match 100.0%; Score 22; DB 14; Length 180882;
Best Local Similarity 100.0%; Pred. No. 2.3; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0; Gaps 0;
QY 1 ATACCTTCAACTCTCGTCCTT 22
|||||
Db 46096 ATACCTTCAACTCTCGTCCTT 46117
|||||

RESULT 9
AC124010/c
LOCUS AC124010
DEFINITION Mus musculus chromosome UNK clone RP23-119N4, WORKING DRAFT
AC124010
VERSION AC124010.2 GI:22475963
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 182400)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 182400)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (06-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 182400)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT On Aug 25, 2002 this sequence version replaced gi:21327639.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@wustl.edu
----- Project Information -----
Center project name: M_BA0119N04
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-terminator Big Dye; 100%
Assembly program: Phrap; version 0.990319
Consensus quality: 183528 bases at least Q40
Consensus quality: 184210 bases at least Q30
Consensus quality: 184756 bases at least Q20
Insert size: 197000; agarose-fp
Insert size: 196816; sum-of-contigs
Quality coverage: 9.53 in Q20 bases; agarose-fp
Quality coverage: 9.61 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 11868: contig of 11868 bp in length
 * 11869: gap of unknown length
 * 11969: contig of 16938 bp in length
 * 28907: gap of unknown length
 * 29007: contig of 43767 bp in length
 * 72774: gap of unknown length
 * 72874: contig of 52053 bp in length
 * 124927: gap of unknown length
 * 125027: contig of 56876 bp in length
 * 181903: gap of unknown length
 * 182003: contig of 398 bp in length.

FEATURES

source
 1. .182400
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="UNK"
 /clone="RP23-119N4"
 1. .11868
 /note="assembly_name:Contig86"
 11869..11968
 /estimated_length=unknown
 11969..28906
 /note="assembly_name:Contig87"
 28907..29006
 /estimated_length=unknown
 29007..72773
 /note="assembly_name:Contig88"
 72774..72873
 /estimated_length=unknown
 72874..124926
 /note="assembly_name:Contig89"
 124927..125026
 /estimated_length=unknown
 125027..181902
 /note="assembly_name:Contig90"
 181903..182002
 /estimated_length=unknown
 182003..182400
 /note="assembly_name:Contig16"

ORIGIN

Query Match 100.0%; Score 22; DB 14; Length 182400;
 Best Local Similarity 100.0%; Pred. No. 2.3;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTCGTCCTT 22
 |||||
 Db 169317 ATACCTTCAACTCTCGTCCTT 169296

RESULT 10

AL772367
 LOCUS
 DEFINITION Mouse DNA sequence from clone RP23-119N4 on chromosome 2, complete sequence.
 ACCESSION AL772367
 VERSION AL772367.7 GI:22759512
 KEYWORDS HTG.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 185378)
 Wood, J.
 Direct Submission
 Submitted (30-AUG-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

COMMENT

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Sep 9, 2002 this sequence version replaced gi:22265488.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; SW, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-119N4 is from the RPCI-23 Mouse PAC Library constructed by the group of Pieter de Jong.
 For further details see http://www.chori.org/bacpac/home.htm
 VECTOR: pBACe3.6.

FEATURES

source
 1. .185378
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="2"
 /clone="RP23-119N4"
 /clone_lib="RPCI-23"

ORIGIN

Query Match 100.0%; Score 22; DB 9; Length 185378;
 Best Local Similarity 100.0%; Pred. No. 2.3;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTCGTCCTT 22
 |||||
 Db 139510 ATACCTTCAACTCTCGTCCTT 139531

RESULT 11

AC119716/c
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-274116, *** SEQUENCING IN PROGRESS
 ***, 17 unordered pieces.

AC119716
 VERSION AC119716.6 GI:25092366
 KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 221547)
 Muzny, D. Marie., Metzker, M. Lee., Abramzon, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,

Delgado O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, I., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, A., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowicz, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuwa, L., Louised, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokeleneh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L., Puzo, M., Quiror, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, K., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajda, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G., and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 221547)
Worley, K.C.
Direct Submission
Submitted (01-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 221547)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:23611286.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
----- Center project name: GWIC

Center clone name: CH230-274116
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 208013 bases at least Q40
Consensus quality: 211685 bases at least Q30
Consensus quality: 213907 bases at least Q20
Estimated insert size: 199182; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 198201: contig of 198201 bp in length
* 198301: gap of unknown length
* 198302: contig of 1448 bp in length
* 198302: contig of 1448 bp in length
* 199750: gap of unknown length
* 199850: contig of 1266 bp in length
* 201115: gap of unknown length
* 201116: contig of 1123 bp in length
* 202136: gap of unknown length
* 202339: contig of 1062 bp in length
* 202439: gap of unknown length
* 203501: gap of unknown length
* 203601: contig of 1305 bp in length
* 204906: gap of unknown length
* 205006: contig of 1187 bp in length
* 206193: gap of unknown length
* 206293: contig of 1093 bp in length
* 207386: gap of unknown length
* 207485: contig of 1034 bp in length
* 208519: gap of unknown length
* 208520: contig of 1139 bp in length
* 209529: gap of unknown length
* 209759: contig of 1583 bp in length
* 209859: contig of 1583 bp in length
* 211442: gap of unknown length
* 211542: contig of 1335 bp in length
* 212877: contig of 1280 bp in length
* 212977: gap of unknown length
* 214257: contig of 2647 bp in length
* 214357: gap of unknown length
* 217003: contig of 1356 bp in length
* 217103: gap of unknown length
* 218459: contig of 1229 bp in length
* 218559: gap of unknown length
* 218660: contig of 1229 bp in length
* 219789: gap of unknown length
* 219889: contig of 1659 bp in length.
* 219889: contig of 1659 bp in length.

FEATURES
source
1. .221547
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clones="CH230-274116"
1. .1272
/note="wgs end extension
clone end:Sp6"
complement(4416. .5263)
/note="clone boundary
clone end:Sp6
site:
end sequence:BZ266352"
192445. .193553
/note="wgs contig"
195743. .196592
/note="clone boundary
clone end:T7
site:
end_sequence:BZ266349"


```

gap      198202..198301
/estimated_length=unknown
gap      199750..199849
/estimated_length=unknown
gap      201116..201215
/estimated_length=unknown
gap      202339..202438
/estimated_length=unknown
gap      203501..203600
/estimated_length=unknown
gap      204906..205005
/estimated_length=unknown
gap      206193..206292
/estimated_length=unknown
gap      207386..207485
/estimated_length=unknown
gap      208520..208619
/estimated_length=unknown
gap      209759..209858
/estimated_length=unknown
gap      211442..211541
/estimated_length=unknown
gap      212877..212976
/estimated_length=unknown
gap      214257..214356
/estimated_length=unknown
gap      217004..217103
/estimated_length=unknown
gap      218460..218559
/estimated_length=unknown
gap      219789..219888
/estimated_length=unknown

Query Match      100.0%; Score 22; DB 14; Length 221547;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATACCTTCAACTCTGCGTCCTT 22
|||||
Db      214156 ATACCTTCAACTCTGCGTCCTT 214135

RESULT 12
LOCUS      AX003328      22 bp      DNA      linear      PAT 24-AUG-2000
DEFINITION      Sequence 21 from Patent WO9929845.
ACCESSION      AX003328
VERSION      AX003328.1 GI:9927145
KEYWORDS      '
SOURCE      Homo sapiens (human)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE      1
AUTHORS      Angulo-Mora, J.F. and Mauffrey, P.
TITLE      Sequences coding for kin17 protein and their applications
JOURNAL      Patent: WO 9929845-A 21 17-JUN-1999;
ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE
(FR)
FEATURES      Location/Qualifiers
source      1..22
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match      92.7%; Score 20.4; DB 6; Length 22;
Best Local Similarity 95.5%; Pred. No. 23;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 ATACCTTCAACTCTGCGTCCTT 22
|||
Db      914 ATCTCTTCAACTCTGCGTCCTT 893

RESULT 15
LOCUS      AX003311/c
DEFINITION      Sequence 33 from Patent WO9929845.
ACCESSION      AX003340
VERSION      AX003340.1 GI:9927153
KEYWORDS      '
SOURCE      Homo sapiens (human)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE      1
AUTHORS      Angulo-Mora, J.F. and Mauffrey, P.
TITLE      Sequences coding for kin17 protein and their applications
JOURNAL      Patent: WO 9929845-A 33 17-JUN-1999;
ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE
(FR)
FEATURES      Location/Qualifiers
source      1..333
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match      92.7%; Score 20.4; DB 6; Length 333;
Best Local Similarity 95.5%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 ATACCTTCAACTCTGCGTCCTT 22
|||||
Db      305 ATCTCTTCAACTCTGCGTCCTT 284

RESULT 14
LOCUS      AX003310/c
DEFINITION      Sequence 3 from Patent WO9929845.
ACCESSION      AX003310
VERSION      AX003310.1 GI:9927127
KEYWORDS      '
SOURCE      Homo sapiens (human)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE      1
AUTHORS      Angulo-Mora, J.F. and Mauffrey, P.
TITLE      Sequences coding for kin17 protein and their applications
JOURNAL      Patent: WO 9929845-A 3 17-JUN-1999;
ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE
(FR)
FEATURES      Location/Qualifiers
source      1..1002
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match      92.7%; Score 20.4; DB 6; Length 1002;
Best Local Similarity 95.5%; Pred. No. 20;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 ATACCTTCAACTCTGCGTCCTT 22
|||||
Db      914 ATCTCTTCAACTCTGCGTCCTT 893

RESULT 15
LOCUS      AX003311/c
DEFINITION      Sequence 33 from Patent WO9929845.
ACCESSION      AX003340
VERSION      AX003340.1 GI:9927153
KEYWORDS      '
SOURCE      Homo sapiens (human)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE      1
AUTHORS      Angulo-Mora, J.F. and Mauffrey, P.
TITLE      Sequences coding for kin17 protein and their applications
JOURNAL      Patent: WO 9929845-A 33 17-JUN-1999;
ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE
(FR)
FEATURES      Location/Qualifiers
source      1..333
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match      92.7%; Score 20.4; DB 6; Length 333;
Best Local Similarity 95.5%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 ATACCTTCAACTCTGCGTCCTT 22
|||||
Db      305 ATCTCTTCAACTCTGCGTCCTT 284

RESULT 14
LOCUS      AX003310/c
DEFINITION      Sequence 3 from Patent WO9929845.
ACCESSION      AX003310
VERSION      AX003310.1 GI:9927127
KEYWORDS      '
SOURCE      Homo sapiens (human)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE      1
AUTHORS      Angulo-Mora, J.F. and Mauffrey, P.
TITLE      Sequences coding for kin17 protein and their applications
JOURNAL      Patent: WO 9929845-A 3 17-JUN-1999;
ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE
(FR)
FEATURES      Location/Qualifiers
source      1..1002
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match      92.7%; Score 20.4; DB 6; Length 1002;
Best Local Similarity 95.5%; Pred. No. 20;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 ATACCTTCAACTCTGCGTCCTT 22
|||||
Db      914 ATCTCTTCAACTCTGCGTCCTT 893

RESULT 15
LOCUS      AX003311/c
DEFINITION      Sequence 33 from Patent WO9929845.
ACCESSION      AX003340
VERSION      AX003340.1 GI:9927153
KEYWORDS      '
SOURCE      Homo sapiens (human)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE      1
AUTHORS      Angulo-Mora, J.F. and Mauffrey, P.
TITLE      Sequences coding for kin17 protein and their applications
JOURNAL      Patent: WO 9929845-A 33 17-JUN-1999;
ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE
(FR)
FEATURES      Location/Qualifiers
source      1..333
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match      92.7%; Score 20.4; DB 6; Length 333;
Best Local Similarity 95.5%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 ATACCTTCAACTCTGCGTCCTT 22
|||||
Db      305 ATCTCTTCAACTCTGCGTCCTT 284

```


Search completed: November 26, 2005, 19:11:14
Job time : 119.932 secs

This page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 12:48:57 ; Search time 14.1112 Seconds
(without alignments)
10390.557 Million cell updates/sec

Title: US-09-555-529-28

Perfect score: 22

Sequence: 1 ataccttcaactctgcgtcctt 22

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq_21.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1990s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001bs.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002bs.*
- 7: Geneseqn2002bs.*
- 8: Geneseqn2003bs.*
- 9: Geneseqn2003bs.*
- 10: Geneseqn2003bs.*
- 11: Geneseqn2003bs.*
- 12: Geneseqn2004bs.*
- 13: Geneseqn2004bs.*
- 14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	100.0	22	2	AAX85572 PCR prime
2	22	100.0	1102	2	AAX85550 CDNA of a
3	22	100.0	1390	2	AAX85570 CDNA of a
4	22	100.0	1458	2	AAX79936 Murine Ki
5	20.4	92.7	22	22	AAX85569 PCR prime
6	20.4	92.7	398	13	ACF87901 Human SIR
7	20.4	92.7	591	4	AAX84095 Human col
8	20.4	92.7	679	13	ADQ58943 Novel can
9	20.4	92.7	1002	2	AAX85552 Probe der
10	20.4	92.7	1002	2	AAX85551 CDNA of a
11	20.4	92.7	1296	2	AAX85549 CDNA of a
12	18.8	85.5	12332	10	AAD55664 Bovine vi
13	17.8	80.9	1432	13	ADX61400 Plant ful
14	17.8	80.9	1689	10	ACF68507 Photorhab
15	17.8	80.9	1921	13	ADX61280 Plant ful
16	17.8	80.9	110000	10	Continuation (13 o
17	17.8	80.9	182624	10	ACF67367 12
18	17.4	79.1	3882	11	ACL28061 Rice abio
19	17.4	79.1	11673	4	ABL20570 Drosophil

C	20	17.4	79.1	11707	4	ABL26444
	21	17.2	78.2	138	4	AAK79342
	22	17.2	78.2	339	10	ADI21316
C	23	17.2	78.2	825	8	ACA29919
	24	17.2	78.2	1820	10	ABZ224202
	25	17.2	78.2	1888	9	AAU62516
	26	17.2	78.2	2000	6	AAZ36306
	27	17.2	78.2	2173	11	ADL33411
	28	17.2	78.2	2800	10	ABZ24200
	29	17.2	78.2	2818	10	ABZ24203
	30	17.2	78.2	2913	10	ADI21761
	31	17.2	78.2	3814	7	ADS73104
	32	17.2	78.2	3814	7	ADM41958
	33	17.2	78.2	3814	10	ABZ24201
	34	17.2	78.2	3814	12	ADO78126
	35	17.2	78.2	3876	6	AAU40003
C	36	17.2	78.2	111206	11	ACN45152
	37	17	77.3	1858	13	ADX32258
	38	17	77.3	1961	13	ADX55232
	39	17	77.3	1974	13	ADX55182
	40	16.8	76.4	498	9	ACC72960
	41	16.8	76.4	498	12	ADL09602
C	42	16.8	76.4	546	12	ADL11629
	43	16.8	76.4	986	13	ADX60098
C	44	16.8	76.4	1589	6	ABL35042
	45	16.8	76.4	2892	8	ADA69873

ALIGNMENTS

RESULT 1

AAX85572 ID AAX85572 standard; cDNA; 22 BP.

XX AAX85572;

XX 07-SEP-1999 (first entry)

DE PCR primer used to amplify human kin17 CDNA sequences.

XX Human; kin17 protein; cell proliferation; fertility;
XX hyperproliferative disease; protein interaction; curved DNA;
XX HIV replication; HIV integration; repair enzyme; PCR primer; ss.

XX Synthetic.

OS Mus sp.

XX FR2772046-A1.

XX 11-JUN-1999.

XX 09-DEC-1997; 97FR-00015536.

XX 09-DEC-1997; 97FR-00015536.

(COMS) COMMISSARIAT ENERGIE ATOMIQUE.

XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;

XX WPI; 1999-359999/31.

XX New DNA coding for human kin17 protein - useful for controlling cell

XX proliferation or fertility.

XX Example 1; Page 12; 69pp; French.

XX PCR primers AAX85571-72 were used to amplify 1000 base pairs of cDNA
XX encoding a human kin17 protein. The mammalian kin17 protein is useful for
XX preparing a medicament for controlling cell proliferation or for
XX controlling fertility. The medicaments can also be used to treat
XX hyperproliferative diseases. Fragments between amino acids 55 and 235
XX (preferably between amino acids 129 and 228) of a mammalian kin17 protein

CC are useful for regulating the interaction between proteins and curved
 CC DNA. The fragment can be used to block replication of HIV or its
 CC integration into the human genome or to target repair enzymes to curved
 CC DNA sites. Expression vectors for kin17 can be used for controlling cell
 CC proliferation

SQ Sequence 22 BP; 4 A; 8 C; 2 G; 8 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 2; Length 22;
 Best Local Similarity 100.0%; Pred. No. 1.5;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGGTCCTT 22
 |||||
 DB 1 ATACCTTCAACTCTGGTCCTT 22

RESULT 2
 AAX85550/c
 ID AAX85550 standard; cDNA; 1102 BP.

XX

AC AAX85550;

DT 07-SEP-1999 (first entry)

XX cDNA of a gene coding for a mouse deleted kin17 protein.

DE kin17 protein; cell proliferation; fertility; hyperproliferative disease;
 KW protein interaction; curved DNA; HIV replication; HIV integration;
 KW repair enzyme; ss.

XX Mus sp.

OS FR2772046-A1.

PN 11-JUN-1999.

XX 09-DEC-1997; 97FR-00015536.

PF 09-DEC-1997; 97FR-00015536.

XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

PA Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;

PI WPI; 1999-359999/31.

DR New DNA coding for human kin17 protein - useful for controlling cell

PT proliferation or fertility.
 PS Claim 4; Page 31; 69pp; French.

XX The present sequence encodes a mouse kin17 protein with amino acids 129-
 CC 228 deleted. The mammalian kin17 protein is useful for preparing a
 CC medicament for controlling cell proliferation or for controlling
 CC fertility. The medicaments can also be used to treat hyperproliferative
 CC diseases. Fragments between amino acids 55 and 235 (preferably between
 CC amino acids 129 and 228) of a mammalian kin17 protein are useful for
 CC regulating the interaction between proteins and curved DNA. The fragment
 CC can be used to block replication of HIV or its integration into the human
 CC genome or to target repair enzymes to curved DNA sites. Expression
 CC vectors for kin17 can be used for controlling cell proliferation

XX Sequence 1102 BP; 373 A; 205 C; 272 G; 252 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 2; Length 1102;
 Best Local Similarity 100.0%; Pred. No. 2.4;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGGTCCTT 22
 |||||
 DB 857 ATACCTTCAACTCTGGTCCTT 836

RESULT 3
 AAX85570/c

ID AAX85570 standard; cDNA; 1390 BP.

XX

AC AAX85570;

XX 07-SEP-1999 (first entry)

DT cDNA of a gene coding for the murine kin17 protein.

DE Mouse; kin17 protein; cell proliferation; fertility;
 KW hyperproliferative disease; protein interaction; curved DNA;
 KW HIV replication; HIV integration; repair enzyme; ss.

XX Mus sp.

OS FR2772046-A1.

PN 11-JUN-1999.

XX 09-DEC-1997; 97FR-00015536.

PF 09-DEC-1997; 97FR-00015536.

XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

PA Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;

PI WPI; 1999-359999/31.

DR New DNA coding for human kin17 protein - useful for controlling cell

PT proliferation or fertility.

PS Claim 21; Page 35-36; 69pp; French.

XX The present sequence encodes a murine kin17 protein. The mammalian kin17
 CC protein is useful for preparing a medicament for controlling cell
 CC proliferation or for controlling fertility. The medicaments can also be
 CC used to treat hyperproliferative diseases. Fragments between amino acids
 CC 55 and 235 (preferably between amino acids 129 and 228) of a mammalian
 CC kin17 protein are useful for regulating the interaction between proteins
 CC and curved DNA. The fragment can be used to block replication of HIV or
 CC its integration into the human genome or to target repair enzymes to
 CC curved DNA sites. Expression vectors for kin17 can be used for
 CC controlling cell proliferation

XX Sequence 1390 BP; 482 A; 256 C; 353 G; 299 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 2; Length 1390;
 Best Local Similarity 100.0%; Pred. No. 2.4;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGGTCCTT 22
 |||||
 DB 1145 ATACCTTCAACTCTGGTCCTT 1124

RESULT 4
 AAQ79936/c

ID AAQ79936 standard; cDNA; 1458 BP.

XX

AC AAQ79936;

XX 25-MAR-2003 (revised)

DT 06-SEP-1995 (first entry)

XX Murine Kin17 cDNA.

DE chromosomal rearrangement; kin17 protein; SOS DNA repair system; RecA;
 KW genotoxic agent; zinc finger; DNA binding protein; ss.
 XX Mus musculus.

XX FH Key Location/Qualifiers
FT primer_bind complement(1..21)
FT FT /*tag= b
FT FT /note= "Oligo L (AAQ79946) binds to complementary strand"
FT FT 22..1434
FT FT /*tag= 1
FT FT /label= kin17 cDNA
FT FT /note= "nucleotides 1-1414; the genomic DNA contains at
FT FT least two introns within this sequence, see Comments"
FT FT 32..149
FT FT /*tag= c
FT FT /note= "Oligo C (AAQ79938) binding site"
FT FT 46..1221
FT FT /*tag= a
FT FT /product= "Kin17"
FT FT /note= "N's in the sequence denote illegible residues"
FT FT complement(67..86)
FT FT /*tag= d
FT FT /note= "Oligo S (AAQ79947) binds to complementary strand"
FT FT 274..1297
FT FT /*tag= e
FT FT /note= "Oligo D (AAQ79939) binding site"
FT FT complement(339..360)
FT FT /*tag= f
FT FT /note= "Oligo K (AAQ79945) binds to complementary strand"
FT FT 451..1474
FT FT /*tag= g
FT FT /note= "Oligo J (AAQ79944) binding site"
FT FT complement(550..567)
FT FT /*tag= h
FT FT /note= "Oligo E (AAQ79940) binds to complementary strand"
FT FT 802..1825
FT FT /*tag= i
FT FT /note= "Oligo F (AAQ79941) binding site"
FT FT complement(839..862)
FT FT /*tag= j
FT FT /note= "Oligo G (AAQ79942) binds to complementary strand"
FT FT complement(1435..1458)
FT FT /*tag= k
FT FT /note= "Oligo B (AAQ79937) binds to complementary strand"
FT FT
FT FR2706487-A1.
FT 23-DEC-1994.
FT 15-JUN-1993; 93FR-00007171.
FT 15-JUN-1993; 93FR-00007171.
FT (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
FT Angulo-Mora JF, Tissier A, Frelat G, Mauffrey P, Guilly M;
FT WPI; 1995-039031/06.
FT Purified murine kin17 protein prepn. for detecting chromosomal
FT rearrangements - also related antibodies, human and murine DNA, primers,
FT probes and vectors, used to assess damage caused by genotoxic agents.
FT Claim 9; Page 33; 54pp; French.
FT The murine Kin17 protein includes a zinc finger domain (see AAR66766),
FT recognises single- and double-stranded DNA (partic. regions of secondary
FT structure), has apparent mol. wt. 43 kD and is recognised by both anti-
FT kin17 antibodies and antibodies against the RecA protein of E.coli. The
FT kin17 protein is involved in DNA repair; it can be used to monitor
FT chromosomal rearrangements following exposure to genotoxic agents. The
FT kin17 cDNA sequence/AAQ79936 consists of a 1414 nucleotide sequence,
FT flanked by primer binding sites; the genomic kin17 DNA is claimed in
FT which an intron is inserted at position 137 of the 1414 nucleotide cDNA
FT and a second intron is located between nucleotides 339-429 of the cDNA.
FT (Updated on 25-MAR-2003 to correct PN field.)

XX SQ Sequence 1458 BP; 496 A; 272 C; 372 G; 311 T; 0 U; 7 Other;
Query Match 100.0%; Score 22; DB 2; Length 1458;
Best Local Similarity 100.0%; Pred. No. 2.4; Mismatches 0; Gaps 0;
Matches 22; Conservative 0; Indels 0; Gaps 0;
QY 1 ATACCTTCAACTCTGCGTCCTT 22
DB 1190 ATACCTTCAACTCTGCGTCCTT 1169
RESULT 5
AAx85569
ID AAX85569 standard; cDNA; 22 BP.
XX AAX85569;
AC AAX85569;
XX 07-SEP-1999 (first entry)
DT PCR primer and probe used to detect human kin17 nucleic acids.
DE Human; kin17 protein; cell proliferation; fertility; probe;
XX Hyperproliferative disease; protein interaction; curved DNA;
KW HIV replication; HIV integration; repair enzyme; PCR primer; ss.
XX Synthetic.
OS Homo sapiens.
PN FR2772046-A1.
XX 11-JUN-1999.
PD 09-DEC-1997; 97FR-00015536.
XX 09-DEC-1997; 97FR-00015536.
PR (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
PI WPI; 1999-359999/31.
DR New DNA coding for human kin17 protein - useful for controlling cell
XX proliferation or fertility.
XX Claim 7; Page 5; 69pp; French.
XX AAX85553-69 represent PCR primers and probes used to detect nucleic acids
XX encoding human kin17 protein. The mammalian kin17 protein is useful for
XX preparing a medicament for controlling cell proliferation or for
XX controlling fertility. The medicaments can also be used to treat
XX hyperproliferative diseases. Fragments between amino acids 55 and 235
XX (preferably between amino acids 129 and 228) of a mammalian kin17 protein
XX are useful for regulating the interaction between proteins and curved
XX DNA. The fragment can be used to block replication of HIV or its
XX integration into the human genome or to target repair enzymes to curved
XX DNA sites. Expression vectors for kin17 can be used for controlling cell
XX proliferation
SQ Sequence 22 BP; 3 A; 8 C; 2 G; 9 T; 0 U; 0 Other;
Query Match 92.7%; Score 20.4; DB 2; Length 22;
Best Local Similarity 95.5%; Pred. No. 8.7; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATACCTTCAACTCTGCGTCCTT 22
DB 1 ATTCCTTCAACTCTGCGTCCTT 22
RESULT 6
ACF87901

KW toxic response; liver necrosis; fatty liver disease;
KW protein adduct formation; hepatitis; dog; ds.
XX
XX Canis familiaris.
XX
XX WO2004063324-A2.
XX
XX
XX 29-JUN-2004.
XX
XX 05-MAY-2003; 2003WO-US013853.
XX
XX 03-MAY-2002; 2002US-0377240P.
XX
XX (GENE-) GENE LOGIC INC.
XX (PFIZ) PFIZER PROD INC.
XX
XX Diggins JC, Porter M, Wei T;
XX WPI; 2004-561890/54.
XX
XX New isolated nucleic acid molecule, useful for drug screening and
XX toxicity assays or for assessing the impact, including toxicity, of a
XX compound, pharmaceutical agent or environmental pollutant on a cell or
XX living organism.
XX
XX Claim 1; SEQ ID NO 8245; 41pp; English.
XX
XX This invention is related to a novel isolated canine nucleic acid
XX sequences and the construction of canine microarrays containing a
XX significant portion of the canine genome. The isolated canine nucleic
XX acid sequences of the invention may be useful for drug screening and
XX toxicity assays. The invention is therefore useful for assessing the
XX impact, including toxicity, of a compound, pharmaceutical agent or
XX environmental pollutant on a cell or living organism. The methods are
XX useful for detecting genes that are up- or down-regulated in canines in a
XX disease state. The sequences are useful as diagnostic agents or markers
XX to detect a cellular response in a sample individually or as part of a
XX gene expression profile. It is also useful as a target for agents that
XX modulate gene expression or activity. The database is useful for
XX producing electronic Northern blots that allow the user to determine the cell
XX type or tissue in which a given gene is expressed and to allow
XX determination of the abundance or expression level of a given gene in a
XX particular tissue or cell. The methods are useful for determining the
XX similarity of a toxic response to one or more individual compounds. The
XX methods are useful for predicting at least one toxic response or the
XX likelihood that a compound or test agent will induce various specific
XX pathologies such as those of the liver (liver necrosis, fatty liver
XX disease, protein adduct formation or hepatitis), those of the kidney,
XX heart, brain or testes, or other pathologies associated with at least one
XX of the toxins. The methods are also useful for predicting or elucidating
XX the potential cellular pathways influenced, induced or modulated by the
XX compound or test agent due to the similarity of the expression profile
XX compared to the profile induced by a known toxin. The present sequence is
XX that of a canine DNA sequence which was claimed for use during the
XX production of a canine microarray of the invention.
XX
XX Sequence 679 BP; 254 A; 106 C; 127 G; 185 T; 0 U; 7 Other;
SQ
Query Match 92.7%; Score 20.4; DB 13; Length 679;
Best Local Similarity 95.5%; Pred. No. 13;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATACCTTCAACTCTGCGTCCTT 22
Db 413 ATTCCTTCAACTCTGCGTCCTT 392
RESULT 9
AA85552/c
ID AAX85552 standard; cDNA; 1002 BP.
XX
AC AAX85552;
XX
XX
Query Match 92.7%; Score 20.4; DB 13; Length 679;
Best Local Similarity 95.5%; Pred. No. 13;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATACCTTCAACTCTGCGTCCTT 22
Db 413 ATTCCTTCAACTCTGCGTCCTT 392
RESULT 9
AA85552/c
ID AAX85552 standard; cDNA; 1002 BP.
XX
AC AAX85552;
XX
XX
Query Match 92.7%; Score 20.4; DB 2; Length 1002;
Best Local Similarity 95.5%; Pred. No. 13;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATACCTTCAACTCTGCGTCCTT 22
Db 1002 ATTCCTTCAACTCTGCGTCCTT 981
RESULT 10
AAX85551/c
ID AAX85551 standard; cDNA; 1002 BP.
XX
XX
AC AAX85551;
XX
XX
Qy 07-SEP-1999 (first entry)
Db cDNA of a gene coding for a truncated human kin17 protein.
XX
XX kin17 protein; cell proliferation; fertility; hyperproliferative disease;
XX protein interaction; curved DNA; HIV replication; HIV integration;
XX repair enzyme; ss.
XX
XX Homo sapiens.
XX
XX FR2772046-A1.
XX
XX 11-JUN-1999.
XX
XX 09-DEC-1997; 97FR-00015536.
XX
XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
XX
XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
XX WPI; 1999-359999/31.
XX
XX New DNA coding for human kin17 protein - useful for controlling cell
XX proliferation or fertility.
XX
XX Claim 7; Page 32-33; 69pp; French.
XX
XX The present sequence is derived from nucleotides 207-1208 of AAX85549,
XX and is a probe used to isolate human kin17 cDNA. The mammalian kin17
XX protein is useful for preparing a medicament for controlling cell
XX proliferation or for controlling fertility. The medicaments can also be
XX used to treat hyperproliferative diseases. Fragments between amino acids
XX 55 and 235 (preferably between amino acids 139 and 228) of a mammalian
XX kin17 protein are useful for regulating the interaction between proteins
XX and curved DNA. The fragment can be used to block replication of HIV or
XX its integration into the human genome or to target repair enzymes to
XX curved DNA sites. Expression vectors for kin17 can be used for
XX controlling cell proliferation
XX
XX Sequence 1002 BP; 374 A; 179 C; 235 G; 214 T; 0 U; 0 Other;
SQ
Query Match 92.7%; Score 20.4; DB 2; Length 1002;
Best Local Similarity 95.5%; Pred. No. 13;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATACCTTCAACTCTGCGTCCTT 22
Db 1002 ATTCCTTCAACTCTGCGTCCTT 981
RESULT 10
AAX85551/c
ID AAX85551 standard; cDNA; 1002 BP.
XX
XX
AC AAX85551;
XX
XX
Qy 07-SEP-1999 (first entry)
Db cDNA of a gene coding for a truncated human kin17 protein.
XX
XX kin17 protein; cell proliferation; fertility; hyperproliferative disease;
XX protein interaction; curved DNA; HIV replication; HIV integration;
XX repair enzyme; ss.
XX
XX Homo sapiens.
XX
XX FR2772046-A1.
XX
XX 11-JUN-1999.
XX
XX 09-DEC-1997; 97FR-00015536.
XX

Db	7433 ATTCCTTCAACTGTGAGTCCTT 7412
RESULT 13	
ID	ADX61400/c
AC	ADX61400 standard; cDNA; 1432 BP.
XX	ADX61400;
XX	
XX	21-APR-2005 (first entry)
DE	Plant full length inert polynucleotide seqid 32243.
XX	
KW	plant protectant; plant growth regulant; gene therapy; plant;
KW	recombinant DNA construct; physical array; plant breeding marker;
KW	cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW	extreme osmotic condition; pathogen tolerance; pest tolerance;
KW	growth rate; cell cycle pathway; disease resistance;
KW	galactomannan production; lignin production; plant growth regulator;
KW	yield; plant growth; plant development; seed oil; protein yield;
KW	protein content; gene; ss.
XX	
OS	Unidentified.
XX	
FN	US2004034888-A1.
XX	
PD	19-FEB-2004.
XX	
PF	28-APR-2003; 2003US-00425114.
XX	
PR	06-MAY-1999; 99US-00304517.
PR	05-NOV-2001; 2001US-00985678.
XX	
PA	(LIUJ/) LIU J.
PA	(ZHOU/) ZHOU Y.
PA	(KOVA/) KOVALIC D K.
PA	(SCRE/) SCREEN S E.
PA	(TABA/) TABASKA J E.
PA	(CAOY/) CAO Y.
XX	
PI	Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
XX	
DR	WPI; 2004-180133/17.
XX	
PT	New recombinant DNA construct, useful for improving plant tolerance to
PT	cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
FT	pests, for conferring increased resistance to plant disease, or for
PT	improving yield.
XX	
PS	Claim 1; SEQ ID NO 32243; 15pp; English.
XX	
CC	The invention describes a recombinant DNA construct comprising a
CC	polynucleotide consisting of a sequence encoding an amino acid sequence
CC	available in electronic form from the US patent office at
CC	ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC	of the invention are also useful in physical arrays of molecules and as
CC	plant breeding markers. The recombinant DNA construct is useful for
CC	improving plant tolerance to cold, heat, drought, herbicides, extreme
CC	osmotic conditions, pathogens or pests, for manipulating growth rate in
CC	plant cells by modification of the cell cycle pathway, for conferring
CC	increased resistance to plant disease, for producing galactomannan,
CC	lignin or plant growth regulators, for increasing the rate of homologous
CC	recombination in plants, for improving yield by modification of
CC	photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC	or by providing improved plant growth and development under at least one
CC	stress condition or for modifying seed oil or protein yield and/or
CC	content. This sequence represents a plant full length insert
CC	polynucleotide that can be used in the recombinant DNA construct of the
CC	invention.
XX	
SQ	Sequence 1432 BP; 406 A; 294 C; 316 G; 416 T; 0 U; 0 Other;
Query Match	80.9%; Score 17.8; DB 13; Length 1432;
Best Local Similarity	90.5%; Pred. No. 2.4e+02;
Matches	19; Conservative 0; Mismatches 2; Indels 0; Gaps 0

Db	7433 ATTCCTTCAACTGTGAGTCCTT 7412
BEST LOCAL SIMILARITY	90.5%; Pred. No. 2.3e+02;
MATCHES	19; Conservative 0; Mismatches 2; Indels 0; Gaps 0
QY	2 TACCTTCAACTGTGCGTCCTT 22
DB	1218 TTCCCTTCAGCTCTGGTCCTT 1198
RESULT 14	
ACF68507/c	
ID	ACF68507 standard; DNA; 1689 BP.
XX	ACF68507;
XX	
DT	20-NOV-2003 (first entry)
XX	
DE	Phototahabdu luminescens nucleotide sequence #6974.
XX	
KW	Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW	detection; food; gene expression; plant; animal; microorganism; toxin;
KW	antibiotic; biopesticide; virulence factor; disease model; plague;
KW	whooping cough; gene; ds.
XX	
OS	Phototahabdu luminescens.
XX	
PN	WO200294867-A2.
XX	
PD	28-NOV-2002.
XX	
PF	07-FEB-2002; 2002WO-IB003040.
XX	
PR	07-FEB-2001; 2001FR-00001659.
XX	
PA	(INSP) INST PASTEUR.
PA	(CNRS) CNRS CENT NAT RECH SCI.
XX	
PI	Duchaud E, Taourit S, Glaeser P, Frangeul L, Kunst F, Danch

QY 2 TACCTTCAACTCTGGTCCTT 22
DB 1333 TACCTTCAACTCTTTGTCTT 1313
SQ Sequence 1921 BP; 491 A; 455 C; 466 G; 509 T; 0 U; 0 Other;
Query Match 80.9%; Score 17.8; DB 13; Length 1921;
Best Local Similarity 90.5%; Pred. NO. 2.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 TACCTTCAACTCTGGTCCTT 22
DB 1628 TTCCTTCAGCTCTGGTCCTT 1608
Search completed: November 26, 2005, 13:40:56
Job time : 16.1112 secs

RESULT 15
ADX61280/c
ID ADX61280 standard; cDNA; 1921 BP.
XX AC ADX61280;
XX DT 21-APR-2005 (first entry)
XX DE Plant full length insert polynucleotide seqid 32123.
XX KW plant protectant; plant growth regulant; gene therapy; plant;
KW recombinant DNA construct; physical array; plant breeding marker;
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW extreme osmotic condition; pathogen tolerance; pest tolerance;
KW growth rate; cell cycle pathway; disease resistance;
KW galactomannan production; lignin production; plant growth regulator;
KW yield; plant growth; plant development; seed oil; protein yield;
KW protein content; gene; ss.
XX OS Unidentified.
XX PN US2004034888-A1.
XX PD 19-FEB-2004.
XX PF 28-APR-2003; 2003US-00425114.
XX PR 06-MAY-1999; 99US-00304517.
XX PR 05-NOV-2001; 2001US-00985678.
XX PA (LIU/) LIU J.
XX PA (ZHOU/) ZHOU Y.
XX PA (KOVA/) KOVALIC D K.
XX PA (SCRE/) SCREEN S E.
XX PA (TABAS/) TABASKA J E.
XX PA (CAOY/) CAO Y.
XX PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
XX WPI; 2004-180133/17.
XX PT New recombinant DNA construct, useful for improving plant tolerance to
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT pests, for conferring increased resistance to plant disease, or for
PT improving yield.
XX PS Claim 1; SEQ ID NO 32123; 15pp; English.
XX CC The invention describes a recombinant DNA construct comprising a
CC polynucleotide consisting of a sequence encoding an amino acid sequence
CC available in electronic form from the US patent office at
CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC increased resistance to plant disease, for producing galactomannan,
CC lignin or plant growth regulators, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This sequence represents a plant full length insert
CC polynucleotide that can be used in the recombinant DNA construct of the
CC invention.
XX

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: November 26, 2005, 12:58:53 ; Search time 122.977 Seconds
(without alignments)
8370.011 Million cell updates/sec

Title: US-09-555-529-28
Perfect score: 22
Sequence: 1 atacctcaactctgcgtcctt 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: gb_est1: *
2: gb_est2: *
3: gb_est3: *
4: gb_hic: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_est7: *
9: gb_gss1: *
10: gb_gss2: *
11: gb_gss3: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	22	100.0	238	1	AV244074
C 2	22	100.0	272	1	AV091267
C 3	22	100.0	283	7	CK334767
4	22	100.0	322	1	A1462821
5	22	100.0	391	1	AU041571
C 6	22	100.0	399	5	BY395138
C 7	22	100.0	430	6	CF617307
C 8	22	100.0	447	2	BB749127
C 9	22	100.0	457	1	AK045702
10	22	100.0	466	2	B2448684
11	22	100.0	512	3	BM228037
12	22	100.0	512	3	BM229452
13	22	100.0	513	2	BE225867
14	22	100.0	560	2	BG072126
15	22	100.0	601	3	BP774644
16	22	100.0	677	2	BE321724
17	22	100.0	1081	2	BE283156
C 18	22	100.0	1154	10	AV412520
C 19	20.4	92.7	202	3	BQ307881
C 20	20.4	92.7	238	2	BF456707
C 21	20.4	92.7	242	6	CB243501
C 22	20.4	92.7	244	1	AV239958

C 23	20.4	92.7	259	6	CF526711
C 24	20.4	92.7	328	2	BE091046
C 25	20.4	92.7	348	1	AI261989
26	20.4	92.7	384	1	AA770446
27	20.4	92.7	386	3	BM661987
28	20.4	92.7	398	8	H75516
C 29	20.4	92.7	437	2	BB749400
C 30	20.4	92.7	469	1	AI087818
31	20.4	92.7	476	1	AI378396
32	20.4	92.7	485	1	AI089251
C 33	20.4	92.7	488	3	BM689918
C 34	20.4	92.7	512	5	BG548883
C 35	20.4	92.7	516	2	BG944189
36	20.4	92.7	521	6	BE349628
C 37	20.4	92.7	543	6	CB158644
38	20.4	92.7	558	5	BU737321
C 39	20.4	92.7	575	3	BM750112
C 40	20.4	92.7	591	1	AJ397270
C 41	20.4	92.7	592	8	BE896845
42	20.4	92.7	604	9	AQ342366
43	20.4	92.7	605	6	CB242699
44	20.4	92.7	606	1	AL597250
C 45	20.4	92.7	660	1	AL597250

ALIGNMENTS

RESULT 1
AV244074/c
LOCUS
DEFINITION
musculus cDNA clone 4831427J19 3' similar to X58472 Mouse KIN17
mRNA for kin17 protein, mRNA sequence.

AV244074 238 bp mRNA linear EST 04-NOV-1999
musculus cDNA clone 4831427J19 3' similar to X58472 Mouse KIN17
mRNA for kin17 protein, mRNA sequence.

AV244074 GI:6231533
EST.
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidae; Muridae; Murinae; Mus.
1 (bases 1 to 238)
Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T.,
Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F.,
Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I.,
Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M.,
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K.,
Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y.,
Suzuki, H., Suzuki, H., Takahashi, F., Tateno, M., Tomihata, N.,
Tsunoda, Y., Watanabe, S., Watanabe, S., Yamamura, T., Yasunishi, A.,
Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Konno, H., et al. 1999)
Unpublished (1999)

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gscc.riken.jp, URL: http://genome.gscc.riken.jp/
Sasaki, N., Izawa, M., Watanabe, M., Ozawa, K., Tanaka, T., Yoneda, Y.,
Matsuyama, S., Carninci, P., Muramatsu, M., Okazaki, Y. and
Hayashizaki, Y.
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M.,
Okazaki, Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)

Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES

source

1. .238
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="4831427J19"
/sex="mixed"
/tissue_type="head"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 0 day neonate head"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'
GAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCGAGTTAATTAATACCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

ORIGIN

Query Match 100.0%; Score 22; DB 1; Length 238;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATACCTTCAACTCTGCGTCCTT 22
|||||
Db 106 ATACCTTCAACTCTGCGTCCTT 85

RESULT 2

AV091267/c
LOCUS 272 bp mRNA linear EST 28-JUN-1999
DEFINITION AV091267 Mus musculus tongue C57BL/6J adult Mus musculus cDNA clone 2310061E15, mRNA sequence.

ACCESSION AV091267

VERSION AV091267

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 272)

REFERENCE

AUTHORS

Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K., Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Niitsu, H., Oda, H., Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tateo, M., Tomaru, Y., Tomimaga, N., Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
RIKEN Mouse ESTs
Unpublished (1999)

TITLE

JOURNAL

COMMENT

Contact: Chie Owa
Genome Science Laboratory
RIKEN

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098

Email: genome-res@rtc.riken.go.jp
Thermotabilization and thermoactivation of thermostable enzymes by trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES

source

Location/Qualifiers
1. .272
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="2310061E15"
/sex="male"
/tissue_type="tongue"
/dev_stage="adult"
/clone_lib="Mus musculus tongue C57BL/6J adult"

ORIGIN

Query Match 100.0%; Score 22; DB 1; Length 272;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATACCTTCAACTCTGCGTCCTT 22
|||||

Db 142 ATACCTTCAACTCTGCGTCCTT 121

RESULT 3

LOCUS

CK334767 283 bp mRNA linear EST 22-DEC-2003
H3107C09-3 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone H3107C09 3', mRNA sequence.

ACCESSION CK334767

VERSION CK334767.1 GI:40290380

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 283)

Tanaka, T.S., Jaradat, S.A., Lim, M.K., Kargul, G.J., Wang, X., Grabovac, M.J., Pantano, S., Sano, Y., Piao, Y., Nagaraja, R., Doi, H., Wood, W.H. III, Becker, K.G. and Ko, M.S.H.
Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray

Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)

JOURNAL

PURVED

COMMENT

Contact: Dawood B. Dudekula

Laboratory of Genetics

National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdna@gsun.grc.nia.nih.gov

Plate: H3107 row: C column: 09

Seq primer: -21M13 Forward

High quality sequence stop: 283

POLYA=Yes.

FEATURES

source

Location/Qualifiers
1. .283
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="niaEST:H3107C09-3"
/db_xref="taxon:10090"
/clone="H3107C09"
/sex="Clones arrayed from a variety of cDNA libraries"
/dev_stage="Clones arrayed from a variety of cDNA libraries"

```

libraries"
/lab_host="DH108"
/clone_lib="NIA Mouse 15K cDNA Clone Set"
/notes="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
clone is among a rearranged set of 15,247 clones from 11
embryo cDNA libraries (including preimplantation stage
embryos from unfertilized egg to blastocyst, embryonic
part of E7.5 embryos, extraembryonic part of E7.5
embryos, and E12.5 female mesonephros/gonad) and one
newborn ovary cDNA library. Average insert size 1.5 kb.
All source libraries are cloned unidirectionally with
Oligo(dT)-Not primers. References include: (1)
Genome-wide expression profiling of mid-gestation
placenta and embryo using a 15,000 mouse developmental
cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97:
9127-9132; (2) Large-scale cDNA analysis reveals phased
gene expression patterns during preimplantation mouse
development, 2000, Development, 127: 1737-1749; (3)
Genome-wide mapping of unselected transcripts from
extraembryonic tissue of 7.5-day mouse embryos reveals
enrichment in the t-complex and under-representation on
the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

ORIGIN
Query Match 100.0%; Score 22; DB 7; Length 283;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATACCTTCAACTCTGCGTCCTT 22
|||||
Db 215 ATACCTTCAACTCTGCGTCCTT 236

RESULT 4
AI462821
LOCUS AI462821 322 bp mRNA linear EST 09-MAR-1999
DEFINITION v848h11.x1 Soares mouse lymph node NbMLN Mus musculus cDNA clone
IMAGE:752229 3', mRNA sequence.
ACCESSION AI462821
VERSION 1 GI:4316851
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 322)
Marra M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
Person, B., Waller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Contact: Maria M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MG1:461213
This clone was previously sequenced on the 5' end only, this new
data is from the 3' end
High quality sequence stop: 315.
Location/Qualifiers
1. :322
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:752229"

FEATURES
source
Query Match 100.0%; Score 22; DB 7; Length 283;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATACCTTCAACTCTGCGTCCTT 22
|||||
Db 215 ATACCTTCAACTCTGCGTCCTT 236

RESULT 6
BY395138/c

libraries"
/lab_host="DH108"
/clone_lib="NIA Mouse 15K cDNA Clone Set"
/notes="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
clone is among a rearranged set of 15,247 clones from 11
embryo cDNA libraries (including preimplantation stage
embryos from unfertilized egg to blastocyst, embryonic
part of E7.5 embryos, extraembryonic part of E7.5
embryos, and E12.5 female mesonephros/gonad) and one
newborn ovary cDNA library. Average insert size 1.5 kb.
All source libraries are cloned unidirectionally with
Oligo(dT)-Not primers. References include: (1)
Genome-wide expression profiling of mid-gestation
placenta and embryo using a 15,000 mouse developmental
cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97:
9127-9132; (2) Large-scale cDNA analysis reveals phased
gene expression patterns during preimplantation mouse
development, 2000, Development, 127: 1737-1749; (3)
Genome-wide mapping of unselected transcripts from
extraembryonic tissue of 7.5-day mouse embryos reveals
enrichment in the t-complex and under-representation on
the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

ORIGIN
Query Match 100.0%; Score 22; DB 1; Length 322;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATACCTTCAACTCTGCGTCCTT 22
|||||
Db 210 ATACCTTCAACTCTGCGTCCTT 231

RESULT 5
AU041571
LOCUS AU041571 391 bp mRNA linear EST 04-DEC-1998
DEFINITION AU041571 Mouse four-cell-embryo cDNA Mus musculus cDNA clone
J1006B10 3', mRNA sequence.
ACCESSION AU041571
VERSION AU041571.1 GI:3955806
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
1 (bases 1 to 391)
Ko, M.S.H., Kitchen, J.R., Wang, X., Threat, T.A., Sun, T.,
Depalma, G.E., Liang, Y., Kargul, G.J., Sharara, R., Lim, M.K. and
Doi, H.
Systematic analyses of genes expressed in 4-cell mouse embryo (The
ERATO/Doi Project at Wayne State University)
Unpublished (1998)
Contact: Hirofumi Doi
Doi Biosymmetry Project, ERATO
Japan Science and Technology Corporation (JST)
Wag Marine East 12F, 2-6 Nakase, Mihama-ku, Chiba 261-71, Japan
Email: hdo@bio.jst.go.jp.
Location/Qualifiers
1. :391
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="J1006B10"
/dev_stage="four-cell-embryo"
/clone_lib="Mouse four-cell-embryo cDNA"

FEATURES
source
Query Match 100.0%; Score 22; DB 1; Length 391;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATACCTTCAACTCTGCGTCCTT 22
|||||
Db 215 ATACCTTCAACTCTGCGTCCTT 236

RESULT 6
BY395138/c

```

LOCUS	BY395138	399 bp	mRNA	linear	EST 12-DEC-2002
DEFINITION	BY395138 RIKEN full-length enriched, CRL-1722 L5178Y-R Mus musculus cDNA clone I730004D10 3', mRNA sequence.				
ACCESSION	BY395138				
VERSION	BY395138.1 GI:26624626				
KEYWORDS	EST.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidae; Muridae; Murinae; Mus.				
	1 (bases 1 to 399)				
AUTHORS	Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Oeato,N., Saito,R., Saito,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Kiyosawa,H., Yagi,K., K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierzki,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Perte,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlstedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,M., Carninci,P., Hayatsu,N., Hirozane-Shisikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukawa,S., Hara,A., Haghizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,D., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.				
	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs				
TITLE	Nature 420, 563-573 (2002)				
JOURNAL	1246851				
PUBLISHED	Contact: Yoshihide Hayashizaki				
COMMENT	Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/ Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission Human Genome Sequences Mamm. Genome. 12, 673-677 (2001) Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.				
FEATURES	Please visit our web site (http://genome.gsc.riken.go.jp) for further details.				
	Location/Qualifiers 1..399 /organism="Mus musculus" /mol_type="mRNA" /strain="DBA/2" /db_xref="taxon:10090" /clone="I730004D10" /cell_line="CRL-1722 L5178Y-R" /clone_lib="RIKEN full-length enriched, CRL-1722 L5178Y-R"				
ORIGIN	Query Match 100.0%; Score 22; DB 5; Length 399; Best Local Similarity 100.0%; P: 0; Mismatches 0; Indels 0; Gaps 0; Matches 22; Conservative 0;				
	QY 1 ATACCTTCAACTCTCGTCCTT 22 DB 185 ATACCTTCAACTCTCGTCCTT 164 				
RESULT 7	CF617307/c				
	LOCUS				
DEFINITION	CF617307				
	AGENCOURT_15774329 NIH_MGC_204 Mus musculus cDNA clone IMAGE:30526444 5', mRNA sequence.				
ACCESSION	CF617307				
	VERSION				
KEYWORDS	CF617307.1 GI:37234283				
	EST.				
SOURCE	Mus musculus (house mouse)				
	Mus musculus				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidae; Muridae; Murinae; Mus.				
	1 (bases 1 to 430)				
REFERENCE	NIH-MGC http://mgc.nci.nih.gov/				
	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)				
AUTHORS	Contact: Daniela S. Gerhard, Ph.D.				
	Office of Cancer Genomics				
TITLE	Bldg. 31 Rm10A07 Bethesda, MD 20892				
	Email: cgapbs-remail.nih.gov				
JOURNAL	Tissue Procurement: Naryan Bhat				
	cDNA Library Preparation: Express Genomics				
COMMENT	DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)				
	DNA Sequencing by: Agencourt Bioscience Corporation				
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:				
	http://image.llnl.gov				
FEATURES	Plate: NDAM609 row: d column: 05				
	High quality sequence stop: 430.				
FEATURES	Location/Qualifiers				
	1..430 /organism="Mus musculus" /mol_type="mRNA" /db_xref="taxon:10090" /clone="IMAGE:30526444" /lab_host="DH10B (phage-resistant)" /clone_lib="NIH_MGC_204" /note="Organ: Placenta; Vector: pExpress-1; Site 1: EcorV; Site 2: NotI; RNA obtained from three placentas from female C57/BL6 mouse at 16 days pregnancy. Tissues were snap-frozen and kept at -80C for two days before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer: 5'-pCAGTACTGTAGTCGCGAGCGGCC(T)25-3' and cloned into the EcorV/NotI sites of pExpress-1. Size-selection >0.75kb resulted in an average insert size of 1.1 kb. This primary, nanoquantity library is normalized to Cots (non-normalized primary library is NIH MGC 223) and was constructed by Express Genomics (Frederick, MD)."				

ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 430;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGGCTCCTT 22
 |||||
 DB 185 ATACCTTCAACTCTGGCTCCTT 164

RESULT 8

BB749127/c
 LOCUS BB749127 RIKEN full-length enriched, pooled tissues, pituitary,
 DEFINITION etc. Mus musculus cDNA clone G030102N09 3', mRNA sequence.

ACCESSION BB749127

VERSION BB749127.1 GI:16153363

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 447)
 Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
 Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K.,
 Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
 Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,
 Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
 Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
 Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
 Tanaka,T., Tomaru,A., Toyota,T., Watanishi,A., Yasunishi,A.,
 Muramatsu,M. and Hayashizaki,Y.
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
 2001)

Unpublished (2001)

JOURNAL

COMMENT Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216

Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
 Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
 Watanishi,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
 Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
 and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,

Sugahara,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.

e mouse tissues.

FEATURES

source

1. :447
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="G030102N09"
 /clone_lib="RIKEN full-length enriched, pooled tissues,

pituitary, etc."

/note="pooled tissues; (tissue_type=pituitary gland,
 dev_stage=adult, sex=male), (tissue_type=thymus,
 dev_stage=adult, sex=male), (tissue_type=forelimb,
 dev_stage=13 days embryo, sex=mixed), (tissue_type=testis,
 dev_stage=13 days embryo, sex=male),
 (tissue_type=olfactory brain, dev_stage=adult, sex=male)"

ORIGIN

Query Match 100.0%; Score 22; DB 2; Length 447;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGGCTCCTT 22
 |||||
 DB 234 ATACCTTCAACTCTGGCTCCTT 213

RESULT 9

AW045702

LOCUS AW045702

DEFINITION UI-M-BH1-akq-d-04-0-UI.e1 NIH BMAP M_S2 Mus musculus cDNA clone

UI-M-BH1-akq-d-04-0-UI 3', mRNA sequence.

ACCESSION AW045702

VERSION AW045702.1 GI:5906231

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 457)

Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

JOURNAL 889548

COMMENT

Contact: Chin, H

National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA

Tel: 301 443 1706

Fax: 301 443 9890

Email: m85t@mail.nih.gov

The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to identify it as a clone from the
 normalized corpus striatum library cDNA library preparation: M.B.
 Soares Lab Clone distribution: NIH BMAP cDNA clones will be made
 available by the means that is soon to be determined. When NIH
 determines the means for distribution of the BMAP cDNA clones, this
 record will be updated accordingly when that means is determined.
 The following repetitive elements were found in this cDNA sequence:

1-21. >AT richlow complexity

Seq primer: M13 Forward

POLYA=yes.

FEATURES

source

1. :457
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UI-M-BH1-akq-d-04-0-UI"
 /dev_stage="27-32 days"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="NIH BMAP M_S2"
 /note="vector: pT73D-Pac (pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; The
 NIH BMAP M_S2 library is a subtracted library derived from
 NIH BMAP M_S1, which in turn is a subtracted library
 derived from a mixture of normalized libraries from ten

regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus). The driver used for subtraction consisted of a pool of 5,000 clones from the NIH BMAP M_S1 library and a pool of 2,000 clones obtained from non-normalized and normalized mouse brain spinal cord libraries.
TAG_TISSUE=corpus-striatum
TAG_LIB=NIH_BMAP_M_S2
TAG_SEQ=ACGGC"

ORIGIN
Query Match 100.0%; Score 22; DB 1; Length 457;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGCGTCCTT 22
|||||
Db 151 ATACCTTCAACTCTGCGTCCTT 172

RESULT 10
BE448684/c
LOCUS
DEFINITION BE448684 466 bp mRNA linear EST 25-JUL-2000
5' similar to TR:O60870 O60870 KIN17 PROTEIN. 1; mRNA sequence.
ACCESSION BE448684
VERSION BE448684.1 GI:9448261
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 466)
TITLE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
NATIONAL Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

UNpublished (1997)
Other ESTs: ut48b03.x1
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:1075249
Seq primer: -40RP from Gibco
High quality sequence stop: 465.

FEATURES
Location/Qualifiers
1..466
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3331085"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/clone_lib="Soares mouse 3NBMS"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCATCTGAAGTGGAGCGCGCTGTTTTTTTTTTTTTTTTTTT (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Bertrand Jordan. Library went through three rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN
Query Match 100.0%; Score 22; DB 2; Length 466;
Best Local Similarity 100.0%; Pred. No. 12;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATACCTTCAACTCTGCGTCCTT 22
|||||
Db 327 ATACCTTCAACTCTGCGTCCTT 306

RESULT 11
BM228037
LOCUS
DEFINITION BM228037 512 bp mRNA linear EST 07-JUN-2003
K0254C06-3 NIA Mouse Unfertilized Egg cDNA Library (Long) Mus musculus cDNA clone NIA:K0254C06 IMAGE:30049853 3', mRNA sequence.
ACCESSION BM228037
VERSION BM228037.2 GI:31485640
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 512)
TITLE Piao, Y., Ko, N.T., Lim, M.K. and KO, M.S.H.
Construction of long-transcript enriched cDNA libraries from submicrogram amounts of total RNAs by a universal PCR amplification method

JOURNAL Genome Res. 11 (9), 1553-1558 (2001)
PUBMED 11544199
COMMENT On Dec 14, 2001 this sequence version replaced gi:17790538.
Other ESTs: K0254C06-5N
Contact: Dawood B. Dudekula
Laboratory of Genetics

National Institute on Aging/National Institutes of Health
333 Casell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@gsun.grc.nia.nih.gov
Plate: K0254 row: C column: 06
Seq primer: -21M13 Forward
High quality sequence stop: 512
POLYA=yes.

FEATURES
Location/Qualifiers
1..512
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="niaEST:K0254C06-3"
/db_xref="taxon:10090"
/clone="NIA:K0254C06 IMAGE:30049853"
/tissue_type="Unfertilized Egg"
/lab_host="DH10B"
/clone_lib="NIA Mouse Unfertilized Egg cDNA Library (Long)"
/note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (<http://lgsun.grc.nia.nih.gov/cDNA>). This is a long-transcript enriched cDNA library [Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]]. Total RNAs were extracted from a pool of 1488 unfertilized eggs.
Double-stranded cDNAs were synthesized with an Oligo (dT) primer [Invitrogen]:
5'-PGACTAGTTCAGATCGGAGCGCGCTTTTTTTTTTTT-3', treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to Lona-linker L1-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.5 kb. The library was constructed by Yulan Piao (NIA)."

ORIGIN

Query Match 100.0%; Score 22; DB 3; Length 512;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGGCTCCTT 22
 |||||
 Db 134 ATACCTTCAACTCTGGCTCCTT 155

RESULT 12

BM229452
 LOCUS 512 bp mRNA linear EST 07-JUN-2003
 DEFINITION Musc musculus cDNA clone NIA:K0278H05 IMAGE:30052216 3', mRNA sequence.

ACCESSION BM229452

VERSION 2 GI:31486939

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 512)

Piao,Y., Ko,N.T., Lim,M.K. and Ko,M.S.H.

Construction of long-transcript enriched cDNA libraries from
 submicrogram amounts of total RNAs by a universal PCR amplification
 method

Genome Res. 11 (9), 1553-1558 (2001)

11544199

On Dec 14, 2001 this sequence version replaced gi:17792386.

Other ESTs: K0278H05-5N

Contact: Dawood B. Dudekula

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdaelg@un.gcr.nia.nih.gov

Plate: K0278 row: H column: 05

Seq primer: -21M13 Forward

High quality sequence stop: 512

POLYA=Yes.

FEATURES

source

Location/Qualifiers

1..512

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="niaEST:K0278H05-3"

/db_xref="taxon:10090"

/clones="NIA:K0278H05 IMAGE:30052216"

/tissue_type="Unfertilized Egg"

/lab_host="DH10B"

/clone_lib="NIA Mouse Unfertilized Egg cDNA Library

(Long)"

/note="vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2:

NotI; Mouse cDNA project by the Laboratory of Genetics,

National Institute on Aging (NIA), Intramural Research

Program, NIH (<http://lgsun.grc.nia.nih.gov/cDNA>). This is

a long-transcript enriched cDNA library (Ref. Genome Res.

11: 1553-1558 (2001). [PMID: 11544199]. Total RNAs were

extracted from a pool of 1488 unfertilized eggs.

Double-stranded cDNAs were synthesized with an Oligo(dT)

primer [Invitrogen:

5'-pGAGTAGTCTAGATCGAGCGCCGCTTTTTTTTTTTT-3'],

treated with T4 DNA polymerase, and purified by

ethanol-precipitation. The cDNAs were ligated to

long-linker Lp-Sal4, purified by phenol/chloroform, and

separated from free linkers by Centricon 100. Then, the

cDNAs were amplified by long-range high fidelity PCR using

Ex Taq polymerase (Takara) with a primer Sal4-S. The

products were purified by phenol/chloroform and Centricon

100. The cDNAs were digested with SalI and NotI enzymes

and cloned into SalI/NotI site of pSPORT1 plasmid vector.

The DH10B E. coli host was transformed with the ligation
 mixture by the standard chemical method. The average
 insert size is about 2.5 kb. The library was constructed
 by Yulan Piao (NIA)."

ORIGIN

Query Match 100.0%; Score 22; DB 3; Length 512;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGGCTCCTT 22
 |||||
 Db 134 ATACCTTCAACTCTGGCTCCTT 155

RESULT 13

BE225867

LOCUS

DEFINITION

513 bp mRNA linear EST 13-MAR-2002
 1a20f09.x1 Mouse E10 5 12 5 Pancreas cDNA Library Mus musculus cDNA
 clone IMAGE:563889 3' similar to TR:060870 O60870 KIN17 PROTEIN.

RNA sequence.

BE225867

VERSION BE225867.1 GI:8931103

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 513)

Melton,D., Meadows,A., Clifton,S., Hillier,L., Marra,M., Pape,D.,

Wyllie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B.,

Ritter,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M.,

McCann,R., Cole,R., Tsagareishvili,R., Williams,T., Jackson,Y. and

Bowers,Y.

WashU-Harvard Pancreas EST Project

Unpublished (2000)

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue

Endocrine Pancreas Consortium

Harvard University, Howard Hughes Medical Institute

Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,

MA 02138

Tel: 617-495-1812

Fax: 617-495-8557

Email: dmelton@biohpc.harvard.edu

Library was constructed by Dr. Douglas Melton DNA sequencing by:

Washington University Genome Sequencing Center For information on

obtaining a clone please contact: Adimika Meadows

(meadows@fas.harvard.edu)

MGI:1847913 This sequence now available from the IMAGE consortium,

for clone orders contact: info@image.llnl.gov

Seq primer: -40UP from Gibco

High quality sequence stop: 384.

Location/Qualifiers

FEATURES

source

1..513

/organism="Mus musculus"

/mol_type="mRNA"

/strain="ICR"

/db_xref="taxon:10090"

/clone="IMAGE:563889"

/sex="Both"

/tissue_type="Pancreatic Bud"

/dev_stage="Embryonic day 10.5 and 12.5, mixed"

/lab_host="DH10B"

/clone_lib="Mouse E10 5 12 5 Pancreas cDNA Library"

/notes="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;

Site_2: Sal 1; Library constructed using SuperScript

Plasmid Library kit (Life Technologies). cDNA made by

oligo-dT priming. Size-selected by column fractionation;

average insert size 1.47 Kb. Primary library, unamplified.

cDNA Library Preparation: Guolin Chen."

ORIGIN

```

Query Match      100.0%; Score 22; DB 2; Length 513;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTCGTCCTT 22
|||||
Db 208 ATACCTTCAACTCTCGTCCTT 229

RESULT 14
BG072126          560 bp mRNA linear EST 18-DEC-2003
LOCUS H3107A09-3 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
DEFINITION BG072126
ACCESSION BG072126
VERSION BG072126.2 GI:40072826
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 560)
AUTHORS Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X.,
Grahovac,M.J., Pantano,S., Sano,Y., Piao,Y., Negarajan,R., Doi,H.,
Wood,W.H. III, Becker,K.G. and Ko,M.S.H.
TITLE Genome-wide expression profiling of mid-gestation placenta and
embryo using a 15,000 mouse developmental cDNA microarray
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
PUBMED 10922068
COMMENT On Jan 26, 2001 this sequence version replaced gi:12554695.
Other ESTs: H3107A09-5
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Casell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgun.grc.nia.nih.gov
This clone set has been freely distributed to the community. Please
visit http://lgun.grc.nia.nih.gov/cDNA/15k.html for details.
Plate: H3107 row: A column: 09
Seq primer: -21M13 Forward
High quality sequence stop: 560
POLYA=Yes.

FEATURES             Location/Qualifiers
     source           1..560
     mol_type="mRNA"
     strain="C57BL/6J"
     db_xref="nia:EST:H3107A09-3"
     clone="H3107A09"
     /dev_stage="Clones arrayed from a variety of cDNA libraries"
     /lab_host="DH10B"
     /clone_lib="NIA Mouse 15K cDNA Clone Set"
     /notes="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; This clone is among a rearranged set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos, and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals

```

```

enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

ORIGIN
Query Match      100.0%; Score 22; DB 2; Length 560;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTCGTCCTT 22
|||||
Db 215 ATACCTTCAACTCTCGTCCTT 236

RESULT 15
BP774644          601 bp mRNA linear EST 02-DEC-2004
LOCUS BP774644
DEFINITION BP774644 mouse (C57BL/6) pancreatic islet library with
recombination-based method Mus musculus cDNA clone mif14029 3',
mRNA sequence.
ACCESSION BP774644
VERSION BP774644.1 GI:50233342
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 601)
AUTHORS Nishimura,M., Yokoi,N., Miki,T., Horikawa,Y., Yoshioka,H.,
Takeda,J., Ohara,O. and Seino,S.
TITLE Construction of a multi-functional cDNA library specific for mouse
pancreatic islets and its application to microarray
JOURNAL DNA Res. 11 (5), 315-323 (2004)
PUBMED 15747579
COMMENT Contact: Susumu Seino
Division of Cellular and Molecular Medicine
Kobe University Graduate School of Medicine
7-5-1 Kusunoki-cho, Chuo-ku, Kobe, Hyogo 650-0017, Japan
Tel: 81-78-382-5360
Fax: 81-78-382-5370
Email: seino@med.kobe-u.ac.jp.

FEATURES             Location/Qualifiers
     source           1..601
     organism="Mus musculus"
     mol_type="mRNA"
     strain="C57BL/6"
     db_xref="taxon:10090"
     clone="mif14029"
     /sex="male"
     /tissue_type="pancreatic islet"
     /dev_stage="adult"
     /clone_lib="mouse (C57BL/6) pancreatic islet library with recombination-based method"

ORIGIN
Query Match      100.0%; Score 22; DB 3; Length 601;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTCGTCCTT 22
|||||
Db 244 ATACCTTCAACTCTCGTCCTT 265

Search completed: November 27, 2005, 00:58:00
Job time : 124.977 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:00:02 ; Search time 4.556 Seconds
(without alignments)

Title: US-09-5551-529-28

perfect score:

Sequence: 1 atacctcaactctgcgtcctt 22

Scoring table: IDENTITY NUC

Scoring table: IDENTIFICATION
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs. 888780828 residues

Total number of bits satisfying chosen parameters: 2606114

```
Minimum DB seq length: 0 |
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```
Database : Issued_Patents_NA.*
1 : /cgn216/ptodata/1/ina/1_COMB.seq.*
2 : /cgn216/ptodata/1/ina/5_COMB.seq.*
3 : /cgn216/ptodata/1/ina/6A_COMB.seq.*
4 : /cgn216/ptodata/1/ina/6B_COMB.seq.*
5 : /cgn216/ptodata/1/ina/H_COMB.seq.*
6 : /cgn216/ptodata/1/ina/PCRUS_COMB.seq.*
7 : /cgn216/ptodata/1/ina/PP_COMB.seq.*
8 : /cgn216/ptodata/1/ina/RE_COMB.seq.*
9 : /cgn216/ptodata/1/ina/backfiles1.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB	ID	Description
C 1	18.8	85.5	675	3	US-09-248-796A-2146 Sequence 2146, Ap
C 2	17.2	78.2	601	3	US-09-949-016-145512 Sequence 145512,
C 3	17.2	78.2	601	3	US-09-949-016-145513 Sequence 145513,
C 4	17.2	78.2	2798	3	US-09-949-016-4106 Sequence 4106, Ap
C 5	17.2	78.2	26103	3	US-09-949-016-16841 Sequence 16841, A
C 6	17.2	78.2	89450	3	US-09-949-016-15848 Sequence 15848, A
C 7	16.8	76.4	15543	3	US-09-949-016-17225 Sequence 17225, A
C 8	16.4	74.5	988	2	US-08-684-862-10 Sequence 10, Appl
C 9	16.4	74.5	1333	2	US-08-684-862-9 Sequence 9, Appl
C 10	16.2	73.6	601	3	US-09-949-016-171590 Sequence 171590, A
C 11	16.2	73.6	601	3	US-09-949-016-179033 Sequence 179033,
C 12	16.2	73.6	601	3	US-09-949-016-179034 Sequence 179034,
C 13	16.2	73.6	703	3	US-09-270-767-1498 Sequence 1498, Ap
C 14	16.2	73.6	703	3	US-09-270-767-16780 Sequence 16780, A
C 15	16.2	73.6	3698	3	US-09-576-594-1098 Sequence 1098, Ap
C 16	16.2	73.6	6133	3	US-09-453-702B-15 Sequence 15, Appl
C 17	16.2	73.6	6133	3	US-10-114-170-15 Sequence 15, Appl
C 18	16.2	73.6	11835	3	US-09-949-016-16904 Sequence 16904, A
C 19	16.2	73.6	46885	3	US-09-949-016-13848 Sequence 13848, A
C 20	16.2	73.6	106310	3	US-09-949-016-16366 Sequence 16366, A
C 21	16.2	73.6	536165	3	US-09-114-808-1 Sequence 1, Appl
C 22	16	72.7	27600	3	US-09-949-016-15290 Sequence 15290, A
C 23	15.8	71.8	601	3	US-09-949-016-183894 Sequence 183894,
C 24	15.8	71.8	732	3	US-09-248-796A-6535 Sequence 6535, Ap

ALIGNMENTS

RESULT 1

US-09-248-796A-2146/c
: Sequence 2146, Application US/09248796A

Patent No. 6747137
GENERAL INFORMATION:

APPLICANT: Keith Weinstock et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248.7

;; CURRENT APPLICATION NUMBER: US/05/248,150A
: CURRENT FILING DATE: 1999-02-12

;; CURRENT FILING DATE: 01/11/2011
;; PRIOR APPLICATION NO.: 12/200,000

;; PRIOR APPLICATION NUMBER: 0
:
: PRIOR FILING DATE: 1998-02-

```

; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 2146
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-2146

Query Match      85.5%; Score 18.8; DB 3; Length 675;
Best Local Similarity 90.9%; Pred. No. 17;
Watched 20. Conservative 0. Mismatches 2; Indels 0; Gaps 0;

```

RESULT 2

US-09-949-016-145512/c
: Sequence 145512, Application US/09949016

; Patent No. 6812339

APPLICANT: VENTER, J. Craig et al.

;
; AFFILIANI: VENTIER, S: CL
; TITLE OF INVENTION: POLY

FILE OF INVENTION: WITH HUMAN DISEASE, I

FILE REFERENCE: CL001307

;
; CURRENT APPLICATION NUMBER: US

```
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 145512
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-145512

Query Match
Best Local Similarity 78.2%; Score 17.2; DB 3; Length 601;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGCGTCCTT 22
    ||||| ||||| ||||| |||||
Db 130 ATACCTTCTACTCTGTGTACTT 109

RESULT 3
US-09-949-016-145513/c
; Sequence 145513, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 145513
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-145513

Query Match
Best Local Similarity 78.2%; Score 17.2; DB 3; Length 601;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGCGTCCTT 22
    ||||| ||||| ||||| |||||
Db 415 ATACCTTCTACTCTGTGTACTT 394

RESULT 4
US-09-949-016-4106
; Sequence 4106, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4106
; LENGTH: 2798
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4106

Query Match
Best Local Similarity 78.2%; Score 17.2; DB 3; Length 2798;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGCGTCCTT 22
    ||||| ||||| ||||| |||||
Db 1572 ATACCTTCTACTCTGTGTACTT 1593

RESULT 5
US-09-949-016-16841/c
; Sequence 16841, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16841
; LENGTH: 26103
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(26103)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16841

Query Match
Best Local Similarity 78.2%; Score 17.2; DB 3; Length 26103;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGCGTCCTT 22
    ||||| ||||| ||||| |||||
Db 16895 AAAGCTTCTACTCTGCGTCCTT 16874

RESULT 6
US-09-949-016-15848
; Sequence 15848, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15848
; LENGTH: 89450
; TYPE: DNA
```

```
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(89450)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15848

Query Match          78.2%; Score 17.2; DB 3; Length 89450;
Best Local Similarity 86.4%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ATACCTTCAACTCTGCGTCCTT 22
Db 84421 ATACCTTCTACTCTGTGTACTT 84442

RESULT 7
US-09-949-016-17225
; Sequence 17225, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17225
; LENGTH: 15543
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(15543)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17225

Query Match          76.4%; Score 16.8; DB 3; Length 15543;
Best Local Similarity 90.0%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 ACCTTCAACTCTGCGTCCTT 22
Db 4395 ACCTTCATCTCTGTGTCCTT 4414

RESULT 8
US-08-684-862-10/c
; Sequence 10, Application US/08684862
; Patent No. 5759541
; GENERAL INFORMATION:
; APPLICANT: Bach, Alfred
; APPLICANT: Hillen, Heinz
; APPLICANT: Bialojan, Siegfried
; TITLE OF INVENTION: No. 5759541el Proteins, the Preparation and Use
; FILE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kell & Weinkauff
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
; COMPUTER: IBM AT-compatible, 80286 processor
; OPERATING SYSTEM: MS-DOS version 5.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/684,862
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/361,705
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/966,040
; FILING DATE: 30-DEC-1992
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
; COMPUTER: IBM AT-compatible, 80286 processor
; OPERATING SYSTEM: MS-DOS version 5.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/684,862
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/361,705
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/966,040
; FILING DATE: 30-DEC-1992
; APPLICATION NUMBER: PCT/EP91/01361
; FILING DATE: 19-JUL-1991
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 988 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: Agkistrodon rhodostoma
; FEATURE:
; LOCATION: 197 to 904
; OTHER INFORMATION: the coding region shown in (2)(ix)(B)
; OTHER INFORMATION: codes for the protein of SEQ ID NO: 5
US-08-684-862-10

Query Match          74.5%; Score 16.4; DB 2; Length 988;
Best Local Similarity 94.4%; Pred. No. 2.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATACCTTCAACTCTGCGT 18
Db 126 ATAGCTTCAACTCTGCGT 109

RESULT 9
US-08-684-862-9/c
; Sequence 9, Application US/08684862
; Patent No. 5759541
; GENERAL INFORMATION:
; APPLICANT: Bach, Alfred
; APPLICANT: Hillen, Heinz
; APPLICANT: Bialojan, Siegfried
; TITLE OF INVENTION: No. 5759541el Proteins, the Preparation and Use
; FILE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kell & Weinkauff
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
; COMPUTER: IBM AT-compatible, 80286 processor
; OPERATING SYSTEM: MS-DOS version 5.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/684,862
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/361,705
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/966,040
; FILING DATE: 30-DEC-1992
; COMPUTER READABLE FORM:
```

```

; APPLICATION NUMBER: PCT/EP91/01361
; FILING DATE: 19-JUL-1991
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1333 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; ORIGINAL SOURCE:
; ORGANISM: Agkistrodon rhodostoma
; FEATURE:
; LOCATION: 231 to 935
; OTHER INFORMATION: the coding region shown in (2) (ix) (B)
; OTHER INFORMATION: codes for the protein of SEQ ID NO: 4
US-08-684-862-9

Query Match 74.5%; Score 16.4; DB 2; Length 1333;
Best Local Similarity 94.4%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGCGT 18
Db 157 ATAGCTTCAACTCTGCGT 140

RESULT 10
US-09-949-016-71590
; Sequence 71590, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 71590
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-71590

Query Match 73.6%; Score 16.2; DB 3; Length 601;
Best Local Similarity 85.7%; Pred. No. 2.9e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGCGTCT 21
Db 495 ATGCCTTGAACCTCTGCATCCT 515

RESULT 11
US-09-949-016-179033
; Sequence 179033, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 179033
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-179033

Query Match 73.6%; Score 16.2; DB 3; Length 601;
Best Local Similarity 85.7%; Pred. No. 2.9e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGCGTCT 21
Db 556 ACACCTTCACCTCTGCTTCT 576

RESULT 12
US-09-949-016-179034
; Sequence 179034, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 179034
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-179034

Query Match 73.6%; Score 16.2; DB 3; Length 601;
Best Local Similarity 85.7%; Pred. No. 2.9e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGCGTCT 21
Db 556 ACACCTTCACCTCTGCTTCT 576

RESULT 13
US-09-270-767-1498/c
; Sequence 1498, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1498
; LENGTH: 703
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-1498

```

Query Match 73.6%; Score 16.2; DB 3; Length 703;
 Best Local Similarity 85.7%; Pred. No. 2.9e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 TACCTTCAACTCTGCGTCCTT 22
 Db 589 TGCCTTCAACTTTCGTCCTT 569

RESULT 14

US-09-270-767-16780/c
 ; Sequence 16780, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 16780
 ; LENGTH: 703
 ; TYPE: DNA
 ; ORGANISM: Drosophila melanogaster
 US-09-270-767-16780

Query Match 73.6%; Score 16.2; DB 3; Length 703;
 Best Local Similarity 85.7%; Pred. No. 2.9e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 TACCTTCAACTCTGCGTCCTT 22
 Db 589 TGCCTTCAACTTTCGTCCTT 569

RESULT 15

US-09-976-594-1098/c
 ; Sequence 1098, Application US/09976594
 ; Patent No. 6673549
 ; GENERAL INFORMATION:
 ; APPLICANT: Furness, Michael
 ; APPLICANT: Buchbinder, Jenny
 ; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
 ; FILE REFERENCE: PA-0041 US
 ; CURRENT APPLICATION NUMBER: US/09/976,594
 ; CURRENT FILING DATE: 2001-10-12
 ; PRIOR APPLICATION NUMBER: 60/240,409
 ; PRIOR FILING DATE: 2000-10-12
 ; NUMBER OF SEQ ID NOS: 1143
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 1098
 ; LENGTH: 3698
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: Incyte ID No. 6673549 197159.1
 US-09-976-594-1098

Query Match 73.6%; Score 16.2; DB 3; Length 3698;
 Best Local Similarity 85.7%; Pred. No. 3.7e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 TACCTTCAACTCTGCGTCCTT 22
 Db 2577 TTCCTTCAAAATCTGTCTT 2557

Search completed: November 27, 2005, 01:10:52
 Job time : 6.556 secs

1715 Page Blank (uspo)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:00:41 ; Search time 25,3623 Seconds
(without alignments)
7173.088 Million cell updates/sec

Title: US-09-555-529-28

Perfect score: 22

Sequence: 1 ataccttcaactctgctctt 22

Scoring table: IDENTITY_NUC

Gapop 10,10 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.Main:

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	20.4	92.7	591	5	US-10-106-698-1187
C 2	18.8	85.5	25	7	US-10-719-956-48210
C 3	18.8	85.5	12332	7	US-10-236-542D-1
C 4	17.8	80.9	1432	7	US-10-425-114-32243
C 5	17.8	80.9	1921	7	US-10-425-114-32123
C 6	17.8	80.9	1995	8	US-10-425-115-167414
C 7	17.4	79.1	4301	7	US-10-437-963-11677
C 8	17.4	79.1	11673	10	US-11-097-143-28096
C 9	17.4	79.1	11707	10	US-11-097-143-36907
C 10	17.2	78.2	25	7	US-10-719-956-48209
C 11	17.2	78.2	637	5	US-10-027-632-226571
C 12	17.2	78.2	637	5	US-10-027-632-226572
C 13	17.2	78.2	637	5	US-10-027-632-226573
C 14	17.2	78.2	637	6	US-10-027-632-226571
C 15	17.2	78.2	637	6	US-10-027-632-226572
C 16	17.2	78.2	637	6	US-10-027-632-226573
C 17	17.2	78.2	825	7	US-10-282-122A-17789
C 18	17.2	78.2	2000	7	US-10-380-727-35
C 19	17.2	78.2	3814	5	US-10-102-524-1701
C 20	17.2	78.2	3876	5	US-10-034-934-50
C 21	17.2	78.2	111206	5	US-10-087-192-1957
C 22	17.2	78.2	367378	6	US-10-312-841-1
C 23	17.2	77.3	1858	7	US-10-425-114-15078

24	17.3	1961	7	US-10-425-114-29972	Sequence 29972, A
25	17.3	1974	7	US-10-425-114-29922	Sequence 29922, A
26	17.3	2617	7	US-10-424-599-101374	Sequence 101374, A
C 27	16.8	498	7	US-10-621-901-33	Sequence 33, Appl
C 28	16.8	502	6	US-10-062-674-716	Sequence 716, App
C 29	16.8	546	7	US-10-621-901-2071	Sequence 2071, App
C 30	16.8	648	5	US-10-027-632-276528	Sequence 276528, A
C 31	16.8	648	6	US-10-027-632-276528	Sequence 13619, A
C 32	16.8	715	5	US-10-027-632-13619	Sequence 13620, A
C 33	16.8	715	5	US-10-027-632-13620	Sequence 13621, A
C 34	16.8	715	5	US-10-027-632-13621	Sequence 13619, A
C 35	16.8	715	6	US-10-027-632-13619	Sequence 13619, A
C 36	16.8	715	6	US-10-027-632-13619	Sequence 13621, A
C 37	16.8	715	6	US-10-027-632-13621	Sequence 13621, A
C 38	16.8	986	7	US-10-425-114-30941	Sequence 30941, A
C 39	16.8	1234	8	US-10-425-115-110431	Sequence 110431, A
C 40	16.8	1266	8	US-10-425-115-110433	Sequence 110433, A
C 41	16.8	1542	6	US-10-156-761-4325	Sequence 4325, App
C 42	16.8	1589	3	US-09-866-050A-571	Sequence 571, App
C 43	16.8	1589	3	US-10-152-661-571	Sequence 40577, A
C 44	16.8	3472	7	US-10-437-963-40577	Sequence 37, Appl
C 45	16.8	13017	5	US-10-139-833-37	

ALIGNMENTS

RESULT 1
US-10-106-698-1187/c
; Sequence 1187, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 1187
; LENGTH: 591
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (587)..(587)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-1187

Query Match 92.7% Score 20.4; DB 5; Length 591;
Best Local Similarity 95.5%; Pred. No. 6.5;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 ATACCTTCAACTCTGCGTCTT 22
|||
Db 239 ATTCCTTCAACTCTGCGTCTT 218

RESULT 2
US-10-719-956-48210/c
; Sequence 48210, Application US/10719956
; Publication No. US20040146910A1
; GENERAL INFORMATION:
; APPLICANT: Xue Mei
; TITLE OF INVENTION: Methods of Genetic Analysis of Rat
; FILE REFERENCE: 3527.1
; CURRENT APPLICATION NUMBER: US/10/719,956

Db 1628 TTCTTCAGCTCTGCGTCCTT 1608

RESULT 7

US-10-437-963-11677

Sequence 11677, Application US/10437963

Publication No. US20040123343A1

GENERAL INFORMATION:

APPLICANT: La Rosa, Thomas J.

APPLICANT: Kovalic, David K.

APPLICANT: Zhou, Yihua

APPLICANT: Cao, Yongwei

APPLICANT: Wu, Wei

APPLICANT: Boukharov, Andrey A.

APPLICANT: Barbazuk, Brad

APPLICANT: Li, Ping

TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With

TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

FILE REFERENCE: 38-21(53221)B

CURRENT APPLICATION NUMBER: US/10/437,963

CURRENT FILING DATE: 2003-05-14

NUMBER OF SEQ ID NOS: 204966

SEQ ID NO 11677

LENGTH: 4301

TYPE: DNA

ORGANISM: Oryza sativa

FEATURE:

OTHER INFORMATION: Clone ID: PAT_MRT4530_17879C.1

US-10-437-963-11677

Query Match 79.1%; Score 17.4; DB 7; Length 4301;

Best Local Similarity 94.7%; Pred. No. 2.3e+02;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 ACCTTCAACTCTGCGTCCT 21

Db 763 ACCTTCAACTCTGCGTCCT 781

RESULT 8

US-11-097-143-36906/c

Sequence 28096, Application US/11097143

Publication No. US20050208558A1

GENERAL INFORMATION:

APPLICANT: Venter, J. Craig

APPLICANT: et al.

TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID

TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE

FILE REFERENCE: CL000728

CURRENT APPLICATION NUMBER: US/11/097,143

CURRENT FILING DATE: 2005-04-04

PRIOR APPLICATION NUMBER: 60/157,832

PRIOR FILING DATE: 1999-10-05

PRIOR FILING DATE: 1999-10-19

PRIOR FILING DATE: 1999-10-28

PRIOR FILING DATE: 1999-11-12

PRIOR FILING DATE: 1999-12-28

PRIOR FILING DATE: 2000-01-12

PRIOR FILING DATE: 2000-02-24

PRIOR FILING DATE: 2000-03-23

NUMBER OF SEQ ID NOS: 43008

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 28096

LENGTH: 11673

TYPE: DNA

US-11-097-143-36906

Query Match 79.1%; Score 17.4; DB 10; Length 11707;

Best Local Similarity 94.7%; Pred. No. 2.5e+02;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 CCTTCAACTCTGCGTCCTT 22

Db 4619 CCTTCAACTCTGCGTCCTT 4601

RESULT 9

US-11-097-143-36907/c

Sequence 36907, Application US/11097143

Publication No. US20050208558A1

GENERAL INFORMATION:

APPLICANT: Venter, J. Craig

APPLICANT: et al.

TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID

TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE

FILE REFERENCE: CL000728

CURRENT APPLICATION NUMBER: US/11/097,143

CURRENT FILING DATE: 2005-04-04

PRIOR APPLICATION NUMBER: 60/157,832

PRIOR FILING DATE: 1999-10-05

PRIOR FILING DATE: 1999-10-19

PRIOR FILING DATE: 1999-10-28

PRIOR FILING DATE: 1999-11-12

PRIOR FILING DATE: 1999-12-28

PRIOR FILING DATE: 2000-01-12

PRIOR FILING DATE: 2000-02-24

PRIOR FILING DATE: 2000-03-23

NUMBER OF SEQ ID NOS: 43008

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 36907

LENGTH: 11707

TYPE: DNA

ORGANISM: DROSOPHILA

US-11-097-143-36907

Query Match 79.1%; Score 17.4; DB 10; Length 11707;

Best Local Similarity 94.7%; Pred. No. 2.5e+02;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 CCTTCAACTCTGCGTCCTT 22

Db 4619 CCTTCAACTCTGCGTCCTT 4601

RESULT 10

US-10-719-956-48209/c

Sequence 48209, Application US/10719956

Publication No. US20040146910A1

GENERAL INFORMATION:

APPLICANT: Xue Mei Zhou

TITLE OF INVENTION: Methods of Genetic Analysis of Rat

FILE REFERENCE: 3527.1

CURRENT APPLICATION NUMBER: US/10/719,956

CURRENT FILING DATE: 2003-11-20

PRIOR APPLICATION NUMBER: 60/427,836

PRIOR FILING DATE: 2002-11-20

NUMBER OF SEQ ID NOS: 699466

SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1

SEQ ID NO 48209

ORGANISM: DROSOPHILA

US-11-097-143-28096

Query Match 79.1%; Score 17.4; DB 10; Length 11673;

Best Local Similarity 94.7%; Pred. No. 2.5e+02;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 CCTTCAACTCTGCGTCCTT 22

Db 4619 CCTTCAACTCTGCGTCCTT 4601

RESULT 9

US-11-097-143-36907/c

Sequence 36907, Application US/11097143

Publication No. US20050208558A1

GENERAL INFORMATION:

APPLICANT: Venter, J. Craig

APPLICANT: et al.

TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID

TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE

FILE REFERENCE: CL000728

CURRENT APPLICATION NUMBER: US/11/097,143

CURRENT FILING DATE: 2005-04-04

PRIOR APPLICATION NUMBER: 60/157,832

PRIOR FILING DATE: 1999-10-05

PRIOR FILING DATE: 1999-10-19

PRIOR FILING DATE: 1999-10-28

PRIOR FILING DATE: 1999-11-12

PRIOR FILING DATE: 1999-12-28

PRIOR FILING DATE: 2000-01-12

PRIOR FILING DATE: 2000-02-24

PRIOR FILING DATE: 2000-03-23

NUMBER OF SEQ ID NOS: 43008

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 36907

LENGTH: 11707

TYPE: DNA

ORGANISM: DROSOPHILA

US-11-097-143-36907

Query Match 79.1%; Score 17.4; DB 10; Length 11707;

Best Local Similarity 94.7%; Pred. No. 2.5e+02;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 CCTTCAACTCTGCGTCCTT 22

Db 4619 CCTTCAACTCTGCGTCCTT 4601

RESULT 10

US-10-719-956-48209/c

Sequence 48209, Application US/10719956

Publication No. US20040146910A1

GENERAL INFORMATION:

APPLICANT: Xue Mei Zhou

TITLE OF INVENTION: Methods of Genetic Analysis of Rat

FILE REFERENCE: 3527.1

CURRENT APPLICATION NUMBER: US/10/719,956

CURRENT FILING DATE: 2003-11-20

PRIOR APPLICATION NUMBER: 60/427,836

PRIOR FILING DATE: 2002-11-20

NUMBER OF SEQ ID NOS: 699466

SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1

SEQ ID NO 48209

```
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-10-719-956-48209

Query Match
Best Local Similarity 78.2%; Score 17.2; DB 7; Length 25;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTCGGTCCTT 22
Db ||||| ||||| ||||| |||||
22 ATACCTTCATGTCGGCCCTT 1

RESULT 11
US-10-027-632-226571
; Sequence 226571, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 226571
; LENGTH: 637
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-226571

Query Match
Best Local Similarity 78.2%; Score 17.2; DB 5; Length 637;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTCGGTCCTT 22
Db ||||| ||||| ||||| |||||
410 ACACCTTCACCTCTGGGTCCTT 431

RESULT 12
US-10-027-632-226572
; Sequence 226572, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 226572
; LENGTH: 637
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-226572

Query Match
Best Local Similarity 86.4%; Pred. No. 2.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTCGGTCCTT 22
Db ||||| ||||| ||||| |||||
410 ACACCTTCACCTCTGGGTCCTT 431

RESULT 13
US-10-027-632-226573
; Sequence 226573, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 226573
; LENGTH: 637
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-226573

Query Match
Best Local Similarity 86.4%; Pred. No. 2.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTCGGTCCTT 22
Db ||||| ||||| ||||| |||||
410 ACACCTTCACCTCTGGGTCCTT 431

RESULT 14
US-10-027-632-226571
; Sequence 226571, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
```


This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:41:08 ; Search time 7.76471 Seconds
(without alignments)
421.056 Million cell updates/sec

Title: US-09-555-529-28

Perfect score: 22

Sequence: 1 ataccttcaactctgcgtcctt 22

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3205263 seqs, 74304013 residues

Total number of hits satisfying chosen parameters: 6410526

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.New.*

1: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
2: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq2.*
9: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq3.*
10: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	17.4	79.1	19	8	US-11-101-244-813058
C 2	17.4	79.1	19	9	US-11-101-244-813058
C 3	16.4	74.5	19	8	US-11-101-244-813016
C 4	16.4	74.5	19	9	US-11-101-244-813016
C 5	15.2	69.1	1290	1	US-10-858-730-150
C 6	15.2	69.1	340000	7	US-11-102-978-3
C 7	14.8	67.3	19	8	US-11-101-244-1543865
C 8	14.8	67.3	19	9	US-11-101-244-1543865
C 9	14.6	66.4	2311	9	US-11-083-784-1543865
C 10	14.2	64.5	19	8	US-11-101-244-495567
C 11	14.2	64.5	19	9	US-11-083-784-495567
C 12	14.2	64.5	2468	1	US-10-995-793-72
C 13	14.2	64.5	3038	7	US-11-113-837-3
C 14	14.2	64.5	3038	7	US-11-113-837-3
C 15	14.2	64.5	3824	1	US-10-131-826A-541
C 16	14.2	64.5	26000	1	US-10-949-720-191
C 17	14	63.6	963	1	US-10-793-626-2987
C 18	14	63.6	1656	1	US-10-793-626-1181
C 19	14	63.6	1846	1	US-10-793-626-4403
C 20	14	63.6	2043	1	US-10-467-962B-100
C 21	14	63.6	2223	7	US-11-137-465-28
C 22	14	63.6	3059	1	US-10-793-626-3536
C 23	14	63.6	3109	1	US-10-689-742-85

C 24	14	63.6	3684	1	US-10-858-730-261
C 25	14	63.6	3753	7	US-11-137-465-29
C 26	14	63.6	4158	1	US-10-793-626-3823
C 27	14	63.6	5495	7	US-11-046-456-33
C 28	14	63.6	5495	7	US-11-046-644-33
C 29	14	63.6	38703	7	US-11-052-544-28
C 30	13.8	62.7	19	8	US-11-101-244-50308
C 31	13.8	62.7	19	8	US-11-101-244-50317
C 32	13.8	62.7	19	8	US-11-101-244-50429
C 33	13.8	62.7	19	8	US-11-101-244-50521
C 34	13.8	62.7	19	8	US-11-101-244-50535
C 35	13.8	62.7	19	8	US-11-101-244-50616
C 36	13.8	62.7	19	8	US-11-101-244-50630
C 37	13.8	62.7	19	8	US-11-101-244-50708
C 38	13.8	62.7	19	8	US-11-101-244-50720
C 39	13.8	62.7	19	8	US-11-101-244-50837
C 40	13.8	62.7	19	8	US-11-101-244-781007
C 41	13.8	62.7	19	8	US-11-101-244-781092
C 42	13.8	62.7	19	8	US-11-101-244-781107
C 43	13.8	62.7	19	8	US-11-101-244-781191
C 44	13.8	62.7	19	8	US-11-101-244-781207
C 45	13.8	62.7	19	8	US-11-101-244-781291

ALIGNMENTS

RESULT 1
US-11-101-244-813058/c
; Sequence 813058, Application US/1101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 813058
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-813058

Query Match 79.1%; Score 17.4; DB 8; Length 19;
Best Local Similarity 94.7%; Pred. No. 3.4;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TACCTTCAACTCTGCGTCC 20
| | | | | | | | | | | | | | | | | | | | | |
Db 19 TTCTTCAACTCTGCGTCC 1

RESULT 2
US-11-083-784-813058/c
; Sequence 813058, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William

; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; PRIOR FILING DATE: 2005-03-18
; PRIOR FILING DATE: 2005-03-18
; PRIOR FILING DATE: 2003-11-14
; PRIOR FILING DATE: 2003-11-14
; PRIOR FILING DATE: 2003-09-10
; PRIOR FILING DATE: 2003-09-10
; PRIOR FILING DATE: 2003-09-10
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 813058
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-813058

Query Match 79.1%; Score 17.4; DB 9; Length 19;
Best Local Similarity 94.7%; Pred. No. 3.4;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 TACCTTCAACTCTGCGTCC 20
Db 19 TTCTTCAACTCTGCGTCC 1

RESULT 3

US-11-101-244-813016/c
; Sequence 813016, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 813016
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-813016

Query Match 74.5%; Score 16.4; DB 8; Length 19;
Best Local Similarity 94.4%; Pred. No. 10;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATACCTTCAACTCTGCGT 18
Db 18 ATTCCTTCAACTCTGCGT 1

RESULT 4

US-11-083-784-813016/c
; Sequence 813016, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin

; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; PRIOR FILING DATE: 2005-03-18
; PRIOR FILING DATE: 2005-03-18
; PRIOR FILING DATE: 2003-11-14
; PRIOR FILING DATE: 2003-11-14
; PRIOR FILING DATE: 2003-09-10
; PRIOR FILING DATE: 2003-09-10
; PRIOR FILING DATE: 2003-09-10
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 813016
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-813016

Query Match 74.5%; Score 16.4; DB 9; Length 19;
Best Local Similarity 94.4%; Pred. No. 10;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATACCTTCAACTCTGCGT 18
Db 18 ATTCCTTCAACTCTGCGT 1

RESULT 5

US-10-858-730-150/c
; Sequence 150, Application US/10858730
; Publication No. US20050255568A1
; GENERAL INFORMATION:
; APPLICANT: Bailey, Richard B.
; APPLICANT: Blomquist, Paul
; APPLICANT: Doten, Reed
; APPLICANT: Driggers, Edward M.
; APPLICANT: Madden, Kevin T.
; APPLICANT: O'Leary, Jessica
; APPLICANT: O'Toole, George
; APPLICANT: Trueheart, Joshua
; APPLICANT: Walbridge, Michael J.
; APPLICANT: Yorgey, Peter S.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID
; TITLE OF INVENTION: PRODUCTION
; FILE REFERENCE: 14184-030001
; CURRENT APPLICATION NUMBER: US/10/858,730
; CURRENT FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/475,000
; PRIOR FILING DATE: 2003-05-30
; PRIOR APPLICATION NUMBER: US 60/551,860
; PRIOR FILING DATE: 2004-03-10
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 150
; LENGTH: 1290
; TYPE: DNA
; ORGANISM: Thermobifida fusca
US-10-858-730-150

Query Match 69.1%; Score 15.2; DB 1; Length 1290;
Best Local Similarity 85.0%; Pred. No. 59;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 3 ACCTTCAACTCTGCGTCTT 22
Db 857 ACCTTCAACTCTGCGTCTT 838

RESULT 6

US-11-102-978-3
; Sequence 3, Application US/11102978

Publication No. US20050250142A1
GENERAL INFORMATION:
APPLICANT: University of Utah Technology Transfer Office
APPLICANT: University of Utah Research Foundation
TITLE OF INVENTION: Diagnosis and Treatment of Herpes Simplex Virus Disease
FILE REFERENCE: 0274-5537.1US
CURRENT APPLICATION NUMBER: US/11/102,978
CURRENT FILING DATE: 2005-04-11
PRIOR APPLICATION NUMBER: PCT/US2003/033152
PRIOR FILING DATE: 2003-10-18
PRIOR APPLICATION NUMBER: 60/419,576
PRIOR FILING DATE: 2002-10-18
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn version 3.2
SEQ ID NO 3
LENGTH: 340000
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: exon
LOCATION: (56948)..(57115)
OTHER INFORMATION: C21orf34 exon
FEATURE:
NAME/KEY: misc feature
LOCATION: (80066)..(81089)
OTHER INFORMATION: Gene VDACC2P; voltage-dependent anion channel isoform 2 pseudogene
FEATURE:
NAME/KEY: exon
LOCATION: (167308)..(167438)
OTHER INFORMATION: C21orf34 exon
FEATURE:
NAME/KEY: exon
LOCATION: (216732)..(216833)
OTHER INFORMATION: C21orf34 exon
US-11-102-978-3

Query Match 69.1%; Score 15.2; DB 7; Length 340000;
Best Local Similarity 85.0%; Pred. No. 90;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 2 TACCTTCAACTCTGCGTCT 21
Db 263712 TACATTCACCTTTCCTCT 263731

RESULT 7
US-11-101-244-1543865
Sequence 1543865, Application US/11/101244
Publication No. US20050246794A1
GENERAL INFORMATION:
APPLICANT: Dharmacon, Inc.
APPLICANT: Khvorova, Anastasia
APPLICANT: Reynolds, Angela
APPLICANT: Leake, Devin
APPLICANT: Marshall, William
APPLICANT: Scaringe, Stephen
TITLE OF INVENTION: Functional and Hyperfunctional siRNA
FILE REFERENCE: 13499US
CURRENT APPLICATION NUMBER: US/11/101,244
CURRENT FILING DATE: 2005-04-07
PRIOR APPLICATION NUMBER: 60/502,050
PRIOR FILING DATE: 2003-09-10
PRIOR APPLICATION NUMBER: 60/426,137
PRIOR FILING DATE: 2002-11-14
NUMBER OF SEQ ID NOS: 1591911
SOFTWARE: Proprietary
SEQ ID NO 1543865
LENGTH: 19
TYPE: RNA
ORGANISM: Homo sapiens
US-11-101-244-1543865

Query Match 67.3%; Score 14.8; DB 8; Length 19;
Best Local Similarity 81.0%; Pred. No. 1.2e+02;

Best Local Similarity 50.0%; Pred. No. 62;
Matches 9; Conservative 7; Mismatches 2; Indels 0; Gaps 0;
Qy 5 CTTCACACTCTGCGTCTT 22
Db 1 CUUUAUAAUCUGCGUCCUU 18
RESULT 8
US-11-083-784-1543865
Sequence 1543865, Application US/11/083784
Publication No. US20050245475A1
GENERAL INFORMATION:
APPLICANT: Dharmacon, Inc.
APPLICANT: Khvorova, Anastasia
APPLICANT: Reynolds, Angela
APPLICANT: Leake, Devin
APPLICANT: Marshall, William
APPLICANT: Scaringe, Stephen
TITLE OF INVENTION: Functional and Hyperfunctional siRNA
FILE REFERENCE: 13499US
CURRENT APPLICATION NUMBER: US/11/083,784
CURRENT FILING DATE: 2005-03-18
PRIOR APPLICATION NUMBER: US/10/714,333
PRIOR FILING DATE: 2003-11-14
PRIOR APPLICATION NUMBER: 60/502,050
PRIOR FILING DATE: 2003-09-10
PRIOR APPLICATION NUMBER: 60/426,137
PRIOR FILING DATE: 2002-11-14
NUMBER OF SEQ ID NOS: 1591911
SOFTWARE: Proprietary
SEQ ID NO 1543865
LENGTH: 19
TYPE: RNA
ORGANISM: Homo sapiens
US-11-083-784-1543865

Query Match 67.3%; Score 14.8; DB 9; Length 19;
Best Local Similarity 50.0%; Pred. No. 62;
Matches 9; Conservative 7; Mismatches 2; Indels 0; Gaps 0;
Qy 5 CTTCACACTCTGCGTCTT 22
Db 1 CUUUAUAAUCUGCGUCCUU 18

RESULT 9
US-11-012-762-67
Sequence 67, Application US/11/012762
Publication No. US20050244815A1
GENERAL INFORMATION:
APPLICANT: Georgia State University Research Foundation, Inc.
TITLE OF INVENTION: Compositions and Methods for Viral Resistance Genes
FILE REFERENCE: GSU1.PCT
CURRENT APPLICATION NUMBER: US/11/012,762
CURRENT FILING DATE: 2004-12-15
PRIOR APPLICATION NUMBER: PCT/US03/19300
PRIOR FILING DATE: 2003-06-19
PRIOR APPLICATION NUMBER: US 60/390,046
PRIOR FILING DATE: 2002-06-19
NUMBER OF SEQ ID NOS: 133
SOFTWARE: PatentIn version 3.2
SEQ ID NO 67
LENGTH: 2311
TYPE: DNA
ORGANISM: Rattus norvegicus
FEATURE:
NAME/KEY: CDS
LOCATION: (41)..(2242)
US-11-012-762-67

Query Match 66.4%; Score 14.6; DB 9; Length 2311;
Best Local Similarity 81.0%; Pred. No. 1.2e+02;

Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ATACCTTCAACTCTGCGTCCT 21
 ||||| |||||
Db 1933 ATACCGCAGCTCTGCGTCCT 1953

```

RESULT 10
US-11-101-244-495567/c
; Sequence 495567, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13498US
; CURRENT APPLICATION NUMBER: US/11/101.244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 495567
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-495567

```

```

RESULT 11
US-11-083-784-495567/c
; Sequence 495567, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/456,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 495567
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-495567

```

Best Local Similarity 84.2%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

RESULT 12

US-10-995-793-72/c
; Sequence 72, Application US/10995793
; Publication No. US20050250123A1
; GENERAL INFORMATION:
; APPLICANT: Yang, Ri-Yao
; APPLICANT: Heu, Daniel K.
; APPLICANT: Liu, Fu-Tong
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Reducing Galectin-12 Activity to Reduce Formation of
; TITLE OF INVENTION: Reducing Galectin-12 Activity to Reduce Formation of
; TITLE OF INVENTION: Reducing Galectin-12 Activity to Reduce Formation of
; FILE REFERENCE: 023070-139910US
; CURRENT APPLICATION NUMBER: US/10/995,793
; CURRENT FILING DATE: 2004-11-22
; PRIOR APPLICATION NUMBER: US 60/524,418
; PRIOR FILING DATE: 2003-11-21
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 72
; LENGTH: 2468
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: mouse galectin-12 CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (141)..(1085)
; OTHER INFORMATION: mouse galectin-12
; FEATURE:
; NAME/KEY: repeat region
; LOCATION: (1233)..(1282)
; OTHER INFORMATION: CA repeats
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (2336)..(2340)
; OTHER INFORMATION: AT-rich motif in 3'-UTR
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (2350)..(2354)
; OTHER INFORMATION: AT-rich motif in 3'-UTR
; OTHER INFORMATION: AT-rich motif in 3'-UTR
US-10-995-793-72

Query Match 64.5%; Score 14.2; DB 1; Length 2468;
Best Local Similarity 84.2%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps

QY 3 ACCTTCAACTCTCGTCCT 21
||||| |||||
Db 514 ACCTTCACTCTCGTTCT 496

RESULT 13

US-11-113-837-3
; Sequence 3, Application US/11113837
; Publication No. US20050255513A1
; GENERAL INFORMATION:
; APPLICANT: Akin, Ali R.
; APPLICANT: Bodie, Elizabeth A.
; APPLICANT: Burrow, Shirley
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Turner, Geoffrey
; APPLICANT: Ward, Michael
; TITLE OF INVENTION: Regulatable Growth of Filamentous Fungi
; FILE REFERENCE: GC682-2
; CURRENT APPLICATION NUMBER: US/11/113.837

```

RESULT 13
US-11-113-837-3
; Sequence 3, Application US/11113837
; Publication NO. US20050255513A1
; GENERAL INFORMATION:
; APPLICANT: Akin, Ali R.
; APPLICANT: Bodie, Elizabeth A.
; APPLICANT: Burrow, Shirley
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Turner, Geoffrey
; APPLICANT: Ward, Michael
; TITLE OF INVENTION: Regulatable Growth of Filamentous Fungi
; FILE REFERENCES: GC682-2
; CURRENT APPLICATION NUMBER: US/11/113,837

```

CURRENT FILING DATE: 2005-04-25
PRIOR APPLICATION NUMBER: US 60/276,571
PRIOR FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: US 60/276,618
PRIOR FILING DATE: 2001-03-14
NUMBER OF SEQ ID NOS: 21
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 3038
TYPE: DNA
ORGANISM: Aspergillus nidulans
US-11-113-837-3

Query Match 64.5%; Score 14.2; DB 7; Length 3038;
Best Local Similarity 84.2%; Pred. No. 2e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 TACCTTCAACTCTGCGTCC 20
Db 780 TGCCTTCAACCTGCGGCC 798

RESULT 14
US-11-113-837-14
Sequence 14, Application US/11113837
Publication No. US20050255513A1
GENERAL INFORMATION:
APPLICANT: Akin, Ali R.
APPLICANT: Bodie, Elizabeth A.
APPLICANT: Burrow, Shirley
APPLICANT: Dunn-Coleman, Nigel
APPLICANT: Turner, Geoffrey
APPLICANT: Ward, Michael
TITLE OF INVENTION: Regulatable Growth of Filamentous Fungi
FILE REFERENCE: GC682-2
CURRENT APPLICATION NUMBER: US/11/113,837
CURRENT FILING DATE: 2005-04-25
PRIOR APPLICATION NUMBER: US 60/276,571
PRIOR FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: US 60/276,618
PRIOR FILING DATE: 2001-03-14
NUMBER OF SEQ ID NOS: 21
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14
LENGTH: 3038
TYPE: DNA
ORGANISM: Aspergillus nidulans
US-11-113-837-14

Query Match 64.5%; Score 14.2; DB 7; Length 3038;
Best Local Similarity 84.2%; Pred. No. 2e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 TACCTTCAACTCTGCGTCC 20
Db 780 TGCCTTCAACCTGCGGCC 798

RESULT 15
US-10-131-826A-541
Sequence 541, Application US/10131826A
Publication No. US20050245730A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.

APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C128
CURRENT APPLICATION NUMBER: US/10/131,826A
CURRENT FILING DATE: 2002-04-24
PRIOR APPLICATION NUMBER: 60/049911
PRIOR FILING DATE: 1997-06-18
PRIOR APPLICATION NUMBER: 60/056974
PRIOR FILING DATE: 1997-08-26
PRIOR APPLICATION NUMBER: 60/059113
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059115
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059117
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059122
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059184
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059352
PRIOR FILING DATE: 1997-09-19
PRIOR APPLICATION NUMBER: 60/059588
PRIOR FILING DATE: 1997-09-19
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 541
LENGTH: 3824
TYPE: DNA
ORGANISM: Homo Sapien
US-10-131-826A-541

Query Match 64.5%; Score 14.2; DB 1; Length 3824;
Best Local Similarity 84.2%; Pred. No. 2e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 CCTTCAACTCTGCGTCCCTT 22
Db 1003 CCTTCAACTCTGCGCCTT 1021

Search completed: November 27, 2005, 02:44:32
Job time : 8.76471 secs

2008 0000 0000